

# Non-discriminatory Trade Policies in Panel Structural Gravity Models: Evidence from Monte Carlo Simulations

Richard Sellner

Institute for Advanced Studies, Macroeconomics and Economic Policy, Vienna, Austria

## Correspondence

R. Sellner, Institute for Advanced Studies, Macroeconomics and Economic Policy, Josefstädter Straße 39, AT-1080 Vienna, Austria.

Email: richard.sellner@ihs.ac.at

## Abstract

This paper provides Monte Carlo (MC) simulation evidence on the performance of methods used for identifying the effects of nondiscriminatory trade policy (NDTP) variables in panel structural gravity models. The benchmarked methods include a fixed effect (FE) estimator that utilizes data on intra national trade flows, the bonus-vetus (BV) and the two-stage fixed effect (FE-2S) estimator. The results indicate that only the FE estimates are unbiased and consistent under very general assumptions of the data generating process. The favourable asymptotic properties of the FE estimator unfold as the number of period  $T$  increases.

## JEL CLASSIFICATION

C33, F10, F13

## KEYWORDS

structural gravity model, nondiscriminatory trade policies, Monte Carlo simulation

## 1 | INTRODUCTION

Estimating the effects of trade policy variables on trade flows is a major interest in international economics. Since the seminal paper of Anderson and Van Wincoop (2003), there has been a substantial amount of empirical research involving the econometric estimation of these effects within a structural

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 John Wiley & Sons Ltd

gravity framework (SGM), noting the importance of theory-consistent treatment of the multilateral resistance terms (MRT) to obtain unbiased estimates. Most of the empirical applications on the structural gravity model focused on the estimation of the effects of discriminatory trade policy measures, among them very prominent regional trade agreements (see Anderson & Yotov, 2016; Bergstrand, Larch, & Yotov, 2015; Egger, Francois, Manchin, & Nelson, 2015, for contributions). Estimation is mostly done via Poisson pseudo-maximum likelihood (PPML), including importer and exporter fixed effects in a cross-sectional, and importer–time, exporter–time fixed effects in a panel setting, to control for the MRT.

This identification strategy will, however, fail if the variable of interest lacks variation in both importer and exporter (i.e., importer–time and exporter–time) dimension, since it will be perfectly collinear to one of the fixed effects. As Head and Mayer (2014) mentioned, such variables will include anything that affects a country's propensity to export/import to/from all destinations or sums, averages and differences of country-specific variables. Among those, economists are especially interested in the effects of nondiscriminatory trade policy (NDTP) measures on trade flows and subsequent welfare effects. NDTP measures include most-favored nation tariffs (see Piermartini & Yotov, 2016), export subsidies or promotion (see Lederman, Olarreaga, & Payton, 2010) and trade facilitation (see Beverelli, Neumueller, & Teh, 2015; Hoekman & Nicita, 2011), among others.

Estimation of NDTP in structural gravity models has only by the time of writing been addressed more thoroughly in the literature. Head and Mayer (2014) devoted a subsection in their summary contribution of the gravity model to this topic and Piermartini and Yotov (2016) mentioned the identification problems of the effects on those variables as one of the challenges in empirical structural gravity estimation. Heid, Larch, and Yotov (2017) suggested using data on intra national trade flows to identify the effects of NDTP. Contrary to the MRT captured by the fixed effects, the NDTP measures should not affect intra national trade, introducing bilateral variation. While this identification method is consistent with structural gravity theory and seems very promising, it requires data on domestic trade, that is, goods and services that are nationally produced and consumed. Since data on such flows is often not readily available or published with a substantial time lag, researchers have often resorted to different methods to obtain estimates on their NDTP measures of interest.

In cases where data on intra national trade flows is unavailable, Head and Mayer (2014) and Piermartini and Yotov (2016) suggested to use a two-stage fixed effects identification strategy, in which the fixed effects, obtained from the first stage, are regressed onto the NDTP measure and other country-specific variables. This approach has been frequently applied in empirical research to identify the effects of collinear variables (see, for instance, Anderson & Yotov, 2016; Eaton & Kortum, 2002; Head & Ries, 2008; Melitz, 2008). Another popular method (see, for instance, Bratt, 2017; Portugal-Perez & Wilson, 2012) is the “bonus vetus OLS” introduced by Baier and Bergstrand (2009). Here, the MRT are approximated by the doubly-demeaned trade cost variables, thus avoiding the collinearity problem, since no fixed effects are specified.

Despite the increasing interest and discussion surrounding NDTP in structural gravity models, little is known about the properties of the proposed estimators. In general, there exist only a few Monte Carlo studies that explicitly take into account a data generating process (DGP) that is consistent with structural gravity theory (see, for instance, Baier & Bergstrand, 2009; Egger & Staub, 2016; Head & Mayer, 2014). Since none of these studies focused on NDTP measures or panel models, this paper aims at closing this gap. The benchmarked methods include: (1) fixed effects estimation on intra national data, (2) bonus-vetus and (3) two-stage fixed effects. The latter two can also be estimated without data on intra national trade flows, the case in which these methods are often employed in empirical research. The paper contributes to the existing literature by analyzing the bias and consistency of these estimators proposed in the literature in Monte Carlo experiments based on a DGP that

is consistent with an economic motivation of the structural gravity model. As most empirical studies following Santos Silva and Tenreyro (2006) implement the estimation via PPML, a DGP based on a generalized linear model (GLM) framework will be employed. Furthermore, it is assumed that NDTP measures only affect border-crossing trade flows. The results will be particularly interesting for empirical researchers that have in the past employed one of the methods or are aiming at identifying the effects of NDTP under a structural gravity model in upcoming research.

The main findings of the paper can be summarized in three recommendations for researchers that are interested in estimating the effects of NDTP measures in a theory-consistent SGM. First, we recommend identifying the effects of NDTP by employing a dataset covering domestic trade flows and using the FE approach outlined in Heid et al. (2017). The MC results outlined in this paper show that this estimator is in general unbiased and consistent. Since the other benchmarked methods, the BV and the FE-2S, yielded biased and inconsistent estimates under plausible assumption regarding the DGP, we strongly recommend against their use for estimating the effects of NDTP. Second, the MC results for the recommended FE emphasized the importance of the number of periods  $T$  for the asymptotic properties of the estimates on the NDTP measure. When facing a trade off between the number of countries  $N$  and periods  $T$ , researchers should, consequently, favor the latter over the former for construction their empirical dataset. Third, we demonstrated that estimation via PPML on a DGP with a Gamma variance process may yield biased estimates on the effect of the NDTP measure, also in the larger samples sizes ( $N = 150$  and  $T = 8$ ) considered in the MC experiments. We therefore recommend to additionally report results from a Gamma PML, especially when employing smaller datasets. Model selection may then be based on information criteria, specification tests or prediction errors (see, e.g., Martínez-Zarzoso, 2013; Santos Silva & Tenreyro, 2006).

The remainder of the paper is organized as follows. The next section gives a short introduction to the structural gravity model and discusses the methods that are benchmarked in the Monte Carlo experiments. Section 3 describes the Monte Carlo simulation design, outlines the scenarios analyzed and closes with the results. Section 4 demonstrates the methods on an empirical application. A final section summarizes the results and discusses the implications and limitations.

## 2 | NONDISCRIMINATORY TRADE POLICIES IN GRAVITY MODELS

### 2.1 | Structural gravity model

As in Egger and Staub (2016), the structural gravity model is motivated on grounds of an endowment economy with Armington differentiation. We extend their cross-sectional set-up to a panel framework<sup>1</sup>. At period  $t$  each country  $i = 1, \dots, N$  is endowed with a volume of goods  $H_{it}$  that it sells at a mill price  $p_{it}$  to earn income  $Y_{it} = p_{it}H_{it}$ . As Anderson and Van Wincoop (2003) showed, the import demand of country  $j$  for goods of  $i$  satisfies<sup>2</sup>

$$X_{ijt} = \frac{p_{it}^\alpha \tau_{ijt}^\alpha Y_{jt}}{\sum_{k=1}^N p_{kt}^\alpha \tau_{kjt}^\alpha}, \quad (1)$$

where  $\tau_{ijt} \geq 1$  are the iceberg trade costs and  $1-\alpha$  is the elasticity of substitution (with  $\alpha < 0$ ) under this model<sup>3</sup>. Imposing the market-clearing condition  $\sum_j X_{ijt} = Y_{it}$  to Equation 1 and solving for the equilibrium market price yields

$$p_{it}^{1-\alpha} = \frac{1}{H_{it}} \sum_{j=1}^N \frac{\tau_{ijt}^\alpha P_{jt} H_{jt}}{\sum_{k=1}^N P_{kt}^\alpha \tau_{kjt}^\alpha}, \quad (2)$$

which implicitly determines  $p_{it}$  and  $Y_{it}$ , given  $\alpha$  and the joint distribution of  $\tau_{ijt}$  and  $H_{it}$  (see Egger & Staub, 2016). Substituting this price back into the import demand equation and defining  $P_{jt}^\alpha = \sum_{i=1}^N (p_{it} \tau_{ijt})^\alpha$  as the CES price index results in the well-known structural gravity model

$$X_{ijt} = Y_{it} Y_{jt} \left( \frac{\tau_{ijt}}{\Pi_{it} P_{jt}} \right)^\alpha, \quad (3)$$

$$\Pi_{it}^\alpha = \sum_{k=1}^N \left( \frac{\tau_{ikt}}{P_{kt}} \right)^\alpha Y_{kt}, \quad (4)$$

$$P_{jt}^\alpha = \sum_{k=1}^N \left( \frac{\tau_{kjt}}{\Pi_{kt}} \right)^\alpha Y_{kt}, \quad (5)$$

with  $Y_{it}$  and  $Y_{jt}$  being production and expenditures, whereas for simplicity balanced trade is assumed.

From the definition of the multilateral resistance terms (MRT) given in Equations 4 and 5 it follows, that  $E[\Pi_{it} \tau_{ijt}] \neq 0$ ,  $E[P_{jt} \tau_{ijt}] \neq 0$  and  $E[\Pi_{it} P_{jt}] \neq 0$ . Addressing this correlation structure will be a central feature of the DGP that will be used for the Monte Carlo simulations. To be consistent with a structural gravity model, prices  $p_{it}$  need to be determined endogenously from the structure of the model. Production and expenditure in turn depend on the prices and, along with the trade costs, determine the trade flows. Equations 1 and 2 can therefore be used to generate the endogenous variables  $p_{it}$ ,  $Y_{it}$ , and the structural part<sup>4</sup> of the trade flows, consistent with the economic structure of the model.

## 2.2 | Econometric specification

Specifying the unobserved trade costs  $\tau_{ijt}^\alpha$  and introducing a stochastic term, the model given by Equations 3 to 5 can be re written in a form that can be econometrically estimated. Denoting  $e_{it} = \ln(Y_{it}/\Pi_{it}^\alpha)$ ,  $m_{jt} = \ln(Y_{jt}/P_{jt}^\alpha)$  and specifying the trade costs by a vector of observable variables  $\beta D_{ijt} + \gamma NIP_{jt} + \theta_{ij} = \ln(\tau_{ijt}^\alpha)$ , leads to

$$X_{ijt} = \exp(e_{it} + m_{jt} + \beta D_{ijt} + \gamma NIP_{jt} + \theta_{ij}) \eta_{ijt}. \quad (6)$$

To keep the model simple, the trade costs are specified by time-varying discriminatory variable  $D_{ijt}$ , a time-varying nondiscriminatory import protection measure  $NIP_{jt}$  and a time-invariant directional pairwise trade cost component  $\theta_{ij}$ .

The assumptions about the multiplicative error  $\eta_{ijt}$  will determine which model will be identified. For instance, OLS on a log-linearized version of Equation 6 will be identified if  $E(\ln \eta_{ijt} | e_{it}, m_{jt}, D_{ijt}, NIP_{jt}) = 0$ , that is, there is no dependence between the covariates and the logarithm of the error term. Assuming  $E(\eta_{ijt} | e_{it}, m_{jt}, D_{ijt}, NIP_{jt}) = 1$ , that is, the errors are mean-independent of the covariates, permits identification via a GLM estimator<sup>5</sup>. Santos Silva and Tenreyro (2006) showed that the errors in gravity models of trade will in general be heteroskedastic and dependent on the covariates. OLS estimation on a log-linear version of Equation 6 will then yield biased estimates owing to Jensen's inequality. Furthermore, the presence of zero flows will provoke ad hoc solutions in log-linear models. Since the bulk of the current empirical gravity models are estimated in

multiplicative form, mostly by PPML following the arguments brought in Santos Silva and Tenreiro (2006), the focus of this paper will be on GLM estimation with a logarithmic link function and a Poisson family. GLM estimates will be consistent if the conditional mean function is correctly specified. The efficiency of the estimator will depend on correct specification of the error variance, that is, the linear exponential family of the density chosen.

There are three different approaches towards a theory-consistent estimation of Equation 6 that controls for the structure imposed by  $e_{it}$  and  $m_{jt}$ . The most common procedure is to include  $N^*T$  exporter–period and  $(N-1)^*T$  importer–period fixed effects for  $e_{it}$  and  $m_{jt}$  (see, Feenstra, 2004; Harrigan, 1996). A second option is to apply a structurally iterated estimator (Egger & Staub, 2016; Head & Mayer, 2014). Given a set of starting values for  $e_{it}$  and  $m_{jt}$ , Equation 6 is estimated via GLM, restricting the coefficients on  $e_{it}$  and  $m_{jt}$  to one. The results are used to obtain an estimate of  $\tau_{ijt}^\alpha$ , which is employed to solve for new estimates of  $e_{it}$  and  $m_{jt}$ . This procedure is repeated until convergence. A third option is the quasi-differences estimation, that uses two products of country-pairs to net out  $e_{it}$  and  $m_{jt}$ . This can be done either by using just-identified set of moment conditions (as described in Charbonneau, 2012) or apply a simply ratio-of-ratios estimator as outlined in Head and Mayer (2014) or Egger and Staub (2016). The latter can be implemented by GLM estimation on the sets of transformed variables  $X_{st} = X_{ikt}X_{jlr}/X_{lkr}X_{ijt}$  and  $d_{st} = (d_{ikt} + d_{jlr}) - (d_{lkr} + d_{ijt})$  with  $d$  being a trade cost variable<sup>6</sup>.

### 2.3 | Identification via intra national trade flows

Given the conditional mean is correctly specified, fixed effects, structurally iterated, and quasi-differences estimation will yield consistent estimates for bilaterally varying trade cost variables. If data on intra national trade flows is available, these methods may also be applied to identify the effects of nondiscriminatory trade policies in a structural gravity model. This identification strategy has just recently been introduced by Heid et al. (2017) and the approach is further discussed in Piermartini and Yotov (2016). As Piermartini and Yotov (2016) mention, the intra national dimension turns the monadic nondiscriminatory trade policy variables into dyadic variables. The key identifying assumption is that nondiscriminatory trade policies do not affect intra national trade. Heid et al. (2017) suggested a fixed effects panel model that can be formulated as follows for a cross-sectional setting:

$$X_{ijt} = \exp \left( \lambda_{it}^{FE} + \chi_{jt}^{FE} + \beta^{FE} D_{ijt} + \gamma^{FE} (NIP_{jt} \times INT_{ij}) + \theta_{ij}^{FE} \right) \eta_{ij}^{FE}. \quad (7)$$

The unobserved terms  $e_{it}$  and  $m_{jt}$  are captured by the fixed effects  $\lambda_{it}^{FE}$  and  $\chi_{jt}^{FE}$ . Furthermore, the time-invariant directional pairwise trade costs, such as distance or common language, can be captured by directional pairwise fixed effects  $\theta_{ij}^{FE}$ . The nondiscriminatory trade policy, per definition, affects all trading partners equally, but will not exert an impact on goods and services that are domestically produced and consumed (see Heid et al., 2017), that is  $X_{iit}$ . Thus,  $NIP_{jt}$  is interacted with an international border dummy  $INT_{ij}$  that is one if  $i \neq j$  and zero otherwise. Identification is possible since the MRT will be identified over all periods and country-pairs, including intra national trade, while the nondiscriminatory trade policies can be identified by exploiting the variation between intra national and international flows.

Note that Equation 7 may also be estimated via the structurally iterated and the quasi-differences approach outlined above. However, the fixed effects estimator has several advantages. In contrast to the structurally iterated estimator, the FE estimator imposes no specific structural assumptions on the MRT. The fixed effects will capture the MRT even if some key variables in determining the trade costs are omitted or unobserved. Furthermore, the structurally iterated estimator is more prone to convergence failures than the fixed effects estimator (see Egger & Staub, 2016). Compared with the

quasi-differences approach, the fixed effects estimator is computationally less demanding<sup>7</sup> and leads to lower standard errors (see Egger & Staub, 2016).

A potential disadvantage of the FE estimator may be that, while the coefficients are consistently estimated, the asymptotic variance might be affected by the incidental parameter problem. This problem will in particular be present in cross-sectional analysis. Cameron and Trivedi (2005) showed that in a multiplicative model with one-way fixed effects, the incidental parameter problem can be avoided by using a concentrated likelihood function, but in the case of a two-way fixed effect model the problem would persist. Egger and Staub (2016) showed in their Monte Carlo simulations using cross-sectional data that the incidental parameter problem strongly affects the  $t$  statistics on their trade cost measure. As a solution to the problem for cross-sectional analysis, Jochmans (2017) derived a GMM estimator for two-way multiplicative gravity models. Regarding panel data applications, Fernández-Vala and Weidner (2016) introduced a correction for non-linear models, including a Poisson model, with two fixed effects. Furthermore, Pfaffermayr (2017, 2018) showed that his proposed constrained PPML, that exploits the theoretical SGM constraints instead of including exporter–period and importer–period dummy variables as fixed effects, is not affected by the incidental parameter problem.

Another disadvantage, especially when interested in product-level analysis, is the availability of intra national trade flows. Readily available high quality data sources are still scarce<sup>8</sup>. Researchers are thus often forced to construct their own datasets based on available international trade flow data and national gross output statistics, with the latter data typically covering fewer countries and having longer publication lags. Furthermore, the probability of encountering zero flows will increase with the level detail of product groups. Some countries may not or only very infrequently export or import certain products at all or with specific trading partners. In such cases convergence problems of the estimator may arise or some of the exporter–period, importer–period, or time-invariant directional pairwise fixed effects may not be identified at all.

Finally, when interested in estimating the effects of more than one NDTP variable, that is, a nondiscriminatory importer–period-specific measure of import protection (NIP) and a nondiscriminatory exporter–period-specific export subsidy (NES) measure, the corresponding parameters may not be jointly identified. Heid et al. (2017) demonstrated a case of perfect collinearity of the fixed effects and the NDTP variables if a specific functional dependence between the NDTP measures across countries exists. Oberhofer, Pfaffermayr, and Sellner (2018) showed in their appendix that if the same country–period-specific measure is included as an importer- and exporter-specific NDTP, i.e.  $NIP_{ct} = NES_{ct}$  for  $c, \dots, C$ , or if these two measures are highly correlated, (near) collinearity will occur.

## 2.4 | Identification methods without intra national trade flows

A challenge for empirical estimation occurs if data on intra national flows is not available. This may be the case because they are not published for the countries of interest, or they are reported with a severe time lag and do not match the periods for which the data on nondiscriminatory trade policies is reported. Since in such a case the nondiscriminatory trade policy measures will not vary over either the exporter or importer dimension, they will be perfectly collinear to the exporter–period or importer–period fixed effect and cannot be identified. Furthermore, the collinearity will lead to convergence failure in the structurally iterated estimates, since  $m_{jt}$  and  $NIP_{jt}$  cannot be jointly identified. The quasi-differences procedure is also not applicable, since  $d_{ikt} + d_{ijt} = d_{ikt} + d_{ijt}$  and thus  $d_{st} = 0$  for all  $s$ , thus none of the three methods discussed above will work.

One way to avoid collinearity is to simply construct a new dyadic variable from two monadic variables, taking into account that the functional form avoids collinearity. Examples from the literature include Lee and Park (2007) or Moïsé, Orliac and Minor (2011) that analyzed the effects of



trade facilitation indicators. However, as Head and Mayer (2014) note, most of the dyadic indicators constructed this way may not have a straightforward interpretation. Another ad hoc solution to the problem, noted in Piermartini and Yotov (2016), is to approximate the MRT by remoteness indexes, which has been the common practice in papers prior to Eaton and Kortum (2002) and Anderson and Van Wincoop (2003). Head and Mayer (2014) discussed various frequently used remoteness measures and found that none of them can capture the structure imposed by the theory. In a similar manner Piermartini and Yotov (2016) do not recommend the use of such remoteness measures to substitute MRT or fixed effects.

As a more promising solution to the problem, Piermartini and Yotov (2016) and Head and Mayer (2014) propose a two-stage fixed effects procedure. First the coefficients on the time-varying discriminatory trade cost variables are estimated along with exporter–period, importer–period and directional pairwise fixed effects. Then the respective fixed effects of the first stage are regressed on the NDTP and other country-period-specific variables. This two-step procedure has been frequently used in the literature. For instance, Eaton and Kortum (2001, 2002) used this set-up to identify key parameters of their trade models, Melitz (2008) estimated the effects of the literacy rate and a measure of linguistic diversity on trade, and Gylfason, Martínez-Zarzoso, and Wijkman (2015) adapted this procedure on a panel with time-varying non bilateral data on corruption and democracy. Given its frequent use, this estimator will be included in the Monte Carlo simulations. Both stages will be estimated via GLM:

$$X_{ijt} = \exp \left( \lambda_{it}^{2S} + \chi_{jt}^{2S} + \theta_{ij}^{2S} + \beta^{2S} D_{ijt} \right) \eta_{ijt}^{2S} \quad (8)$$

$$\exp \left( \hat{\lambda}_{jt}^{2S} \right) = \exp \left( \psi^{2S} + \delta^{2S} \ln Y_{jt} + \gamma^{2S} NIP_{jt} \right) v_{jt}^{2S}. \quad (9)$$

The flaws of this procedure for obtaining an estimate of  $\gamma^{2S}$  under fixed effects assumption have been thoroughly outlined in the discussion that followed the fixed effects vector decomposition model introduced in Plümpner and Troeger (2007). As Greene (2011) noted, it is not possible to identify the fixed effects and collinear variables under the assumptions of a fixed effects model. Only the preceding linear mixture of the two is estimable. The parameter  $\gamma^{2S}$  is only separately identified if the fixed effects assumption is violated. Another related approach to identify collinear variables was given in Egger (2005), who extended the Hausman and Taylor (1981) model (HTM) to the gravity framework. Besides being a log-linear model, the HTM imposes strong exogeneity assumptions on the discriminatory variables used as instruments for the identification of the effects of the NDTP measures<sup>9</sup>.

Another commonly applied approach to avoid the collinearity problem is the bonus vetus (BV) method (see Baier & Bergstrand, 2009), that approximates the multilateral resistance terms by a first-order log-linear Taylor series expansion around a symmetric trade cost world. To implement this method, compute  $x_{ijt}^{MRS} = x_{ijt} - MRS(x_{ijt})$  for each trade cost variable  $x$ , with  $MRS(x_{ijt}) = \sum_{k=1}^N x_{kjt} \theta_{kt} + \sum_{l=1}^N x_{ilt} \theta_{lt} - \sum_{k=1}^N \sum_{l=1}^N x_{klt} \theta_{kt} \theta_{lt}$  and  $\theta_{it} = Y_{it} / \sum_{k=1}^N Y_{kt}$ . The normalized trade flows  $X_{ijt} / (Y_{it} Y_{jt})$  are then regressed on the transformed variables  $x_{ijt}^{MRS}$ . A more robust version for estimation purposes is to replace the GDP-weights with equal weights  $1/N$  (see Baier & Bergstrand, 2010), which assumes a symmetric world in trade costs and economic sizes. While the BV method was originally derived for OLS on log-linearized model of Equation 6, BV-transformed variables haven often been used in PPML estimation (see, Bratt, 2017; Hoekman & Nicita, 2011; Moisé & Sorescu, 2013). Given its continuing popularity in empirical applications, the BV method (in its more robust, equally weighted specification) is included in the Monte Carlo simulations:

$$X_{ijt}/(Y_{it}Y_{jt}) = \exp\left(\theta_{ij}^{BV} + \beta^{BV}D_{ijt}^{MRS} + \gamma^{BV}NIP_{jt}^{MRS}\right)\eta_{ijt}^{BV}. \quad (10)$$

In cases of no missing data and assuming a log-linear model with homoskedastic errors, Head and Mayer (2014) showed in Monte Carlo simulations that the BV method yields unbiased estimates for discriminatory trade policy variables in a cross-section. It will also be demonstrated later, that under certain conditions the BV method also yields unbiased estimates on the NDTP variable in a panel setting. The use of the BV method is often motivated on grounds of collinearity in cases where data on intra national flows is missing. It is therefore illustrative to assess the performance of the BV method in a GLM estimation framework without data on intra national trade flows.

Estimates of such a model are subject to several sources of bias. First, under a multiplicative non linear model with heteroskedastic errors, the log-linearized BV-transformations will lead to biased estimates, as shown in Santos Silva and Tenreiro (2006). Since the BV method resembles a double-demeaning procedure, their GLM counterpart would invoke concentrating out the mean effects of exporters and importers or applying some quasi-differences approaches as described above. Such a procedure would strongly complicate the use of this method, whose popularity is a result of its simplicity for application. A second source of bias comes from missing observations, a problem that is usually encountered in empirical applications of the gravity model. As the MC results of Head and Mayer (2014) showed, this bias will be negligible for discriminatory continuous variables. The third source of bias can be attributed to the correlation between the latent inward multilateral resistance term  $m_{jt}$  and the importer–period-specific NDTP variable  $NIP_{jt}$ .

For comparison with the above models, a naive gravity model is added to the MC simulations:

$$X_{ijt} = \exp\left(\theta_{ij}^{naive} + \beta^{naive}D_{ijt} + \gamma^{naive}NIP_{jt} + \omega_1 \ln Y_{it} + \omega_2 \ln Y_{jt}\right)\eta_{ijt}^{naive}. \quad (11)$$

The naive model relates to the basic Newtonian model with exporters and importers GDP's as mass variables. Estimates of  $\gamma$  obtained by the naive gravity model are biased since they omit the MRT that are correlated with the NDTP variable. Prior to the structural gravity model, researchers usually extended the naive gravity model by remoteness indexes. As mentioned before, none of the remoteness variables suggested captures the structure implied by the MRT. In order to avoid selecting an arbitrary remoteness index for an augmented gravity model, the naive gravity model is used as a benchmark.

### 3 | MONTE CARLO SIMULATIONS

Monte Carlo simulation studies based on a DGP that is consistent with structural gravity theory are still scarce. Baier and Bergstrand (2009) employed Monte Carlo simulations to motivate that their bonus vetus OLS method yields results similar to the customized non-linear least squares estimator of Anderson and Van Wincoop (2003). They used observed data on GDP, distance, and borders of the McCallum (1995) Canadian–United States province trade dataset for the simulation analysis, fixing the coefficients on distance and border to a true value. The obtained trade costs were used to solve for the MRT, which were in turn used to construct the expected trade flows of their log-linear model. Prior to estimation they added a log-normal error term with a variance such that a non structural gravity regressions yields an  $R^2$  of between 0.7 to 0.8.

Head and Mayer (2014) adopted a similar approach, using actual data on GDP, distance and regional trade agreements. In contrast to Baier and Bergstrand (2009) they included the log-normal error term in the trade cost, prior to calculation of the MRT, rather than prior to estimation as in Equation 6.



They emphasize this difference in design, noting that they pursued a structural instead of a statistical approach to the error term<sup>10</sup>. The variance of their error term was calibrated to the root mean squared error of a least squares dummy-variable regression on their data. Similar to Baier and Bergstrand (2009), they restrict their analysis to log-linear models and benchmarked a series of linear estimators, including fixed effects, structurally iterated, quasi-differences, and BV.

As the focus of this paper is on GLM estimation, the simulation approaches of Baier and Bergstrand (2009) and Head and Mayer (2014) for log-linear models cannot be applied. The DGP that will be used in the following, is based on the design introduced in Egger and Staub (2016). As in Baier and Bergstrand (2009) their DGP consists of an deterministic structural part that determines the expected trade flows and a stochastic part that introduces noise subsequently.

### 3.1 | Data generating process

The following Monte Carlo experiments are applied to empirical regressors as well as to fully simulated regressors. The DGP is the same for both datasets and obeys the restrictions imposed by the structural gravity model outlined in Section 2. Using empirical data on the regressors and effect sizes of the trade cost variables from the empirical literature has the advantage of producing data properties that are likely to be encountered in practice. In contrast, a full simulation of all regressors enables the researcher to fine tune dispersion and correlation properties of the data and latent parameters.

#### 3.1.1 | Empirical regressors

As empirical regressors we use data on gross domestic product (GDP), the simple mean of the applied most-favored nation tariff rate (MFN) and a dummy variable for regional trade agreements for 150 countries<sup>11</sup> over 3-year intervals between 1994 and 2015. Data on GDP in constant U.S.\$ million and tariffs is obtained from the World Bank Database and data on regional trade agreements is taken from “Mario Larch’s Regional Trade Agreements Database” (see Egger & Larch, 2008). Missing values in the most-favored nation tariff variable were interpolated by the conditional mean prediction from a Poisson regression on the logarithm of GDP, dummy variables indicating the income group definition of the World Bank, a full set of importer country dummies and year dummies<sup>12</sup>. The deterministic part of the trade flows is then generated by setting the linear index of the trade costs to

$$\ln T_{ijt} = \beta RTA_{ijt} + \gamma (\ln(1 + MFN_{jt}) \times INT_{ij}) + \theta_{ij}, \quad (12)$$

solving for the multilateral resistances in Equations 4 and 5 using the empirical measures for GDP and substituting back to Equation 3. The resulting deterministic trade flows will be realized in expectation  $E(X_{ijt}) = \mu_{ijt}$ . This deterministic or structural part is then augmented by a stochastic component, as will be described later. The coefficient on regional trade agreements  $D_{ijt} = RTA_{ijt}$  and the most-favored nation tariff  $NIP_{jt} = \ln(1 + MFN_{jt})$  are set to  $\beta = 0.5$  and  $\gamma = -5$ , which corresponds to values<sup>13</sup> found in the empirical literature (see, Head & Mayer, 2014). Since there is no natural empirical benchmark on how to choose the directional pairwise trade cost components  $\theta_{ij}$ , we draw them from a normal distribution with mean  $-0.6$  and standard deviation of  $0.3$  to resemble key moments of the distribution of  $0.5 \times RTA_{ijt} - 5 \times (\ln(1 + MFN_{jt}) \times INT_{ij})$ , that is, the sum of the linear trade costs index of the other two components. Setting  $\theta_{ij}$  this way specifies equal contributions of observed and unobserved components in expectation.

### 3.1.2 | Fully-simulated regressors

Our second series of MC experiments is based on a fully simulated data set, extending the framework of Egger and Staub (2016) to a panel including an importer-specific NDTP variable and directional pairwise trade cost components. Again, the first part of the DGP is the deterministic component, that obeys the structural gravity theory and will be realized in expectation. For the initial period  $t = 1$ , we draw four correlated variables from a multivariate normal distribution

$$z_{ij}^H, z_{ij}^D, z_{ij}^{NIP}, z_{ij}^\theta \sim \mathcal{MVN}(\mu_z, \Sigma_z) \quad (13)$$

with covariance matrix

$$\Sigma_z = v_z \times \begin{pmatrix} \sigma_H^2 & 0.95 \times \sigma_H \sigma_D & 0.95 \times \sigma_H \sigma_{NIP} & 0.95 \times \sigma_H \sigma_\theta \\ 0.95 \times \sigma_D \sigma_H & \sigma_D^2 & 0.95 \times \sigma_D \sigma_{NIP} & 0.95 \times \sigma_D \sigma_\theta \\ 0.95 \times \sigma_{NIP} \sigma_H & 0.95 \times \sigma_{NIP} \sigma_D & \sigma_{NIP}^2 & 0.95 \times \sigma_{NIP} \sigma_\theta \\ 0.95 \times \sigma_\theta \sigma_H & 0.95 \times \sigma_\theta \sigma_D & 0.95 \times \sigma_\theta \sigma_{NIP} & \sigma_\theta^2 \end{pmatrix}, \quad (14)$$

with the superscripts  $H, D, NIP$  and  $\theta$  denoting the auxiliary variable for endowment, the time-varying discriminatory, the time-varying nondiscriminatory and the directional pairwise trade cost component. Means, standard deviations and correlations are set in a similar manner as in Egger and Staub (2016), with  $\mu_z = (3, -2, 0, -2)$ ,  $\sigma_H = 3\sqrt{N}/4$ ,  $\sigma_D = \sigma_{NIP} = 5$ , and  $\sigma_\theta = 3$ . In the baseline scenario the variance scaling factor  $v_z$  is set to 0.1. The endowments and the trade cost variables for the initial period  $t = 1$  are then obtained by

$$H_{i,t=1} = \exp\left(\sum_{j=1}^N z_{ij}^H / N\right) N, \quad (15)$$

$$D_{ij,t=1} = \left(\frac{\exp(z_{ij}^D)}{1 + \exp(z_{ij}^D)}\right)^{-\alpha/4}, \quad (16)$$

$$NIP_{j,t=1} = \left(1/N \sum_{i=1}^N \frac{\exp(z_{ij}^{NIP})}{1 + \exp(z_{ij}^{NIP})}\right)^{-\alpha/4}, \quad (17)$$

with  $\alpha = -4$  in the baseline scenario, following Anderson and Van Wincoop (2003). The variables are constructed for the full set of  $N^2$  trade flows including intra trade flows. As in Egger and Staub (2016),

$$\theta_{ij} = \left(\frac{\exp(z_{ij}^\theta)}{1 + \exp(z_{ij}^\theta)}\right)^{-\alpha/4}, \quad (18)$$

increasing  $N$  will increase the mean and variance but will not affect the coefficient of variation or the endowment share of a country relative to the world. The variables for the periods  $t > 1$  are generated by applying growth factors drawn from uniform distributions:

$$H_{i,t+1} = H_{i,t} \times g_{H_i}, \quad g_{H_i} \sim \mathcal{U}[0.98; 1.04] \quad (19)$$

$$D_{ij,t+1} = D_{ij,t} \times g_{D_{ij}}, \quad g_{D_{ij}} \sim \mathcal{U}[1;1.02] \quad (20)$$

The total trade cost component  $T_{ijt} = \tau_{ijt}^\alpha$  is then specified by:

$$NIP_{j,t+1} = NIP_{j,t} \times g_{NIP_{j,t}}, \quad g_{NIP_{j,t}} \sim \mathcal{U}[0.98;1]. \quad (21)$$

$$\ln T_{ijt} = \beta D_{ijt} + \gamma (NIP_{jt} \times INT_{ij}) + \theta_{ij}. \quad (22)$$

The coefficients of the DGP are set to  $\beta = 0.5$  and  $\gamma = -5$  to facilitate comparison with the DGP of the empirical dataset described above. The structural part is derived using  $H_{it}$  and  $T_{ijt}$  to solve for  $p_{it}$  in Equation 2. Recognizing that Equation 1 only holds in expectation,  $Y_{it} = p_{it}H_{it}$  and  $T_{ijt} = \tau^\alpha$  are then used to obtain  $E(X_{ijt}) = \mu_{ijt}$ .

In the next step, the observed trade flows—for both the empirical and the fully simulated dataset—are obtained from the structural part by multiplying the deterministic  $\mu_{ijt}$  with the stochastic  $\eta_{ijt}$  component

$$X_{ijt} = \mu_{ijt} \eta_{ijt}. \quad (23)$$

The errors  $\eta_{ijt}$  with  $E(\eta_{ijt}) = 1$  are drawn from the heteroskedastic log-normal distribution

$$\eta_{ijt} = \exp\left(z_{ijt}^\eta\right), \quad z_{ijt}^\eta \sim N(-0.5\sigma_{\eta,ijt}^2, \sigma_{\eta,ijt}^2). \quad (24)$$

The simulation exercises are carried out assuming  $\sigma_{\eta,ijt}^2 = \ln\left(1 + \mu_{ijt}^{-1}\right)$  in the baseline setting. In this setting, GLM with a Poisson variance function, that is, PPML, will lead to an asymptotically efficient estimator. In the Monte Carlo simulations, the deterministic components are derived from the empirical data or constructed by drawing the simulated dataset once. In all but one experiment, the full sample size of  $N = 150$  countries and  $T = 8$  periods is employed for solving the model whereas only the largest  $n = 50, 100, 150$  countries in terms of GDP or endowment and the first  $t = 3, 5, 8$  periods are selected for the respective sample sizes. Within each experiment, the stochastic part is added in each replication as described in Equations 23 and 24.

### 3.2 | Description of MC scenarios

The first two Monte Carlo experiments are conducted on the data generating process using the empirical data. Table 1 shows the data properties of the empirical data generating process. Since the largest  $n$  countries in terms of GDP are considered, the dispersion—given by the coefficient of variation  $CV$ —in trade flows increases with  $N$  and decreases in  $T$ . Also note that the correlation—denoted by  $\rho(\cdot, \cdot)$ —between the discriminatory and nondiscriminatory variable decreases with increasing  $N$  and  $T$ .

In a second experiment, the robustness of the PPML estimator with respect to a different assumption of the variance process is assessed. In a GLM the choice of the linear exponential family distribution will lead to a particular variance function, under which the estimator is asymptotically efficient. In the baseline scenario the errors are drawn from Equation 24 with  $\sigma_{\eta,ij}^2 = \ln\left(1 + \mu_{ij}^{-1}\right)$ , resulting in  $V(X_{ijt}) = \mu_{ijt}$ , that is, a Poisson family. While the PPML is most frequently used in empirical research, other applied PML estimators include Gamma (see, Martínez-Zarzoso, 2013; Santos Silva & Tenreyro, 2006) and negative binomial (see Burger, Van Oort, & Linders, 2009; Head, Mayer, & Ries, 2009). Estimators of these families will be optimal in terms of efficiency if  $\sigma_{\eta,ij}^2 = \ln(2)$  and  $\sigma_{\eta,ij}^2 = \ln\left(2 + \mu_{ij}^{-1}\right)$ , respectively. As a robustness check for the PPML, it will be applied to a DGP with a Gamma variance

**TABLE 1** Data properties of the empirical dataset

	N = 50			N = 100			N = 150		
	T = 3	T = 5	T = 8	T = 3	T = 5	T = 8	T = 3	T = 5	T = 8
$CV(T_{ijt})$	0.49	0.48	0.47	0.50	0.49	0.48	0.51	0.49	0.49
$CV(D_{ijt})$	1.47	1.39	1.29	2.07	1.96	1.79	2.56	2.47	2.24
$CV(NIP_{jt})$	0.78	0.80	0.80	0.68	0.70	0.71	0.59	0.62	0.63
$CV(e_{it})$	0.14	0.15	0.15	0.14	0.15	0.15	0.14	0.15	0.15
$CV(m_{jt})$	0.33	0.31	0.28	0.36	0.33	0.31	0.38	0.35	0.33
$\rho(D_{ijt}, NIP_{jt})$	0.11	0.10	0.09	0.10	0.09	0.05	0.06	0.05	0.01
$\rho(NIP_{jt}, m_{jt})$	-0.86	-0.85	-0.83	-0.87	-0.85	-0.85	-0.88	-0.86	-0.85
$CV(X_{ijt})$	7.28	7.02	6.47	14.11	13.57	12.38	21.11	20.29	18.50
Gamma variance: $\sigma_{n,ijt}^2 = \ln(2)$									
$CV(X_{ijt})$	9.38	9.30	8.80	18.29	17.86	16.76	27.33	26.72	25.07

Note.  $CV(X_{ijt})$  is the mean coefficient of variation over 5,000 replications.

function<sup>14</sup> for the stochastic component, which results in  $V(X_{ijt}) = \mu_{ijt}^2$ . Changes in the error distribution will only affect the data properties via the dispersion in the generated trade flows. Specifying a Gamma family, increases the dispersion in trade flows compared with the baseline, see last row in Table 1.

An additional three MC experiments are conducted on the DGP using the fully simulated dataset. The data properties of the above outlined baseline setting of the fully simulated DGP are given in Panel (a) of Table 2. Data for different  $N$  and  $T$  are generated from the same full sample of  $N = 150$  and  $T = 8$  in each experiment, including the 50, 100, and 150 countries with the largest endowment in the initial period and the 3, 5, and 8 earliest periods. Thus, by construction the dispersion in endowments and trade flows will increase with increasing  $N$ . Note that there is a significant degree of correlation between the nondiscriminatory trade policy variable  $NIP_{jt}$  and the latent inward multilateral resistance term  $m_{jt}$ .

In the first scenario we break this correlation by slightly modifying the DGP described above. In particular,  $NIP_{jt}$  is replaced by a variable  $\widehat{NIP}_{jt}$  drawn from a standard normal distribution such that  $\rho(NIP_{jt}, m_{jt}) = 0^{15}$ . Then we derive  $\beta D_{ijt} = \ln T_{ijt} - \theta_{ij} - \gamma(\widehat{NIP}_{jt} \times INT_{ij})$ . To ensure that  $\rho(\widehat{NIP}_{jt}, m_{jt}) = 0$  for each  $N = 50, 100, 150$  and  $T = 3, 5, 8$ , we generated each sample size separately rather than generating the largest sample size once and then reducing it as in the other experiments. Hence, there is no dependence of the dispersion in endowments and trade flows on the sample size. The modified DGP increases the dispersion of  $D_{ijt}$ , and introduces a high correlation between the  $D_{ijt}$  and  $NIP_{jt}$  (see Panel (b) in Table 2). The coefficient of variation is suppressed since  $NIP_{jt}$  has zero mean in this scenario.

The next two scenarios are taken from Egger and Staub (2016). The third scenario simulates the effects of an increased dispersion in endowments and trade flows, by setting  $v_z = 0.3$ . As shown in Panel (c) of Table 2 this effectively increases the dispersion in all components of the structural part of the DGP compared with the baseline scenario. In the fourth scenario, a higher elasticity of substitution of  $\alpha = -9$  is assumed, which substantially increases the dispersion in the total trade costs  $T_{ijt}$  and trade flows  $X_{ijt}$ , see Panel (d).

Each scenario is simulated for the following four estimators outlined above: (1) fixed effects model (FE), (2) Two-stage fixed effects estimator (FE-2S), (3) bonus-vetus with equal country weights  $1/N$  (BV), and (4) a non-structural gravity model (naive). Models (2) to (4) are estimated on both the full sample covering  $N^2 \times T$  and the reduced sample covering only  $(N^2 - N) \times T$  observations of international trade flows, whereas the fixed effects model (1) as given in Equation 7 is only identified in the full sample with intra national trade flows  $X_{iit}$  included. All methods are estimated by PPML, including the scenario in which the DGP is specified via a GLM with a Gamma variance. In each scenario,  $S = 5,000$  replications of the DGP for a small ( $N = 50$ ), medium ( $N = 100$ ), and large ( $N = 150$ ) country sample size are performed. Furthermore, we vary the time periods over a short ( $T = 3$ ), medium ( $T = 5$ ) and long ( $T = 8$ ) panel.

As the largest panel dataset with  $N = 150$  and  $T = 8$  leads to 180,000 observations with  $N \times T = 1,200$  identified exporter–period fixed effects  $\lambda_{it}$ ,  $(N-1) \times T = 1,192$  identified importer–period fixed effects  $\chi_{jt}$ , and  $N^2 - N = 22,350$  identified directional pairwise fixed effects  $\theta_{ij}$ , we are confronted with high-dimensional fixed effects and computations that are not computationally feasible using standard software for GLM models. We thus employ the algorithms introduced in Stammann (2018) that utilize a weighted version of the Frisch–Waugh–Lovell theorem with alternating projections in a Netwon–Raphson algorithm for GLM models. The routines used are available in the R-package “alpaca” and output all  $N \times T \times 2 + N^2$  fixed effects, that is, including the ones that would be unidentified when using a dummy variable regression framework. Thus, for the two-stage fixed effects estimator on the full sample including intra national observations, we recovered the corresponding

TABLE 2 Data properties of the fully-simulated dataset

	N = 50			N = 100			n = 150		
	T = 3	T = 5	T = 8	T = 3	T = 5	T = 8	T = 3	T = 5	T = 8
Panel (a): Baseline									
$CV(H_t)$	0.16	0.16	0.17	0.19	0.19	0.20	0.25	0.26	0.26
$CV(T_{ijt})$	1.68	1.69	1.70	1.75	1.76	1.77	1.78	1.79	1.80
$CV(D_{ijt})$	-0.59	-0.59	-0.59	-0.57	-0.57	-0.58	-0.57	-0.57	-0.57
$CV(NIP_{jt})$	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.07
$CV(\epsilon_{it})$	0.12	0.12	0.13	0.12	0.12	0.12	0.14	0.14	0.14
$CV(m_{jt})$	0.21	0.21	0.22	0.23	0.23	0.24	0.28	0.28	0.28
$\rho(T_{ijt}, H_t)$	0.02	0.02	0.01	0.04	0.04	0.04	0.06	0.06	0.06
$\rho(D_{ijt}, NIP_{jt})$	-0.03	-0.03	-0.02	-0.03	-0.03	-0.02	-0.04	-0.04	-0.03
$\rho(NIP_{jt}, m_{jt})$	0.48	0.47	0.43	0.48	0.47	0.43	0.40	0.39	0.36
$CV(X_{ijt})$	1.69	1.70	1.71	1.76	1.77	1.79	1.84	1.85	1.86
Panel (b):									
$\rho(NIP_{jt}, m_{jt}) = 0$									
$CV(H_t)$	0.25	0.25	0.24	0.22	0.22	0.22	0.25	0.26	0.26
$CV(T_{ijt})$	1.73	1.77	1.95	1.80	1.85	1.99	1.80	1.86	2.03
$CV(D_{ijt})$	-1.73	-1.73	-1.72	-1.73	-1.73	-1.71	-1.74	-1.74	-1.71
$CV(NIP_{jt})$	—	—	—	—	—	—	—	—	—
$CV(\epsilon_{it})$	0.14	0.14	0.14	0.12	0.13	0.13	0.14	0.14	0.14
$CV(m_{jt})$	0.44	0.45	0.47	0.32	0.34	0.36	0.30	0.32	0.34
$\rho(T_{ijt}, H_t)$	0.11	0.10	0.09	0.07	0.07	0.06	0.06	0.06	0.05
$\rho(D_{ijt}, NIP_{jt})$	0.97	0.97	0.96	0.97	0.97	0.96	0.97	0.97	0.96
$\rho(NIP_{jt}, m_{jt})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$CV(X_{ijt})$	1.70	1.71	1.72	1.82	1.83	1.84	1.84	1.85	1.86

(Continues)



TABLE 2 Continued

	N = 50			N = 100			n = 150		
	T = 3	T = 5	T = 8	T = 3	T = 5	T = 8	T = 3	T = 5	T = 8
Panel (c): $v_z = 0.3$									
$CV(H_{jt})$	0.25	0.25	0.25	0.32	0.32	0.32	0.43	0.43	0.43
$CV(T_{jtr})$	1.70	1.70	1.71	1.72	1.72	1.73	1.76	1.77	1.78
$CV(D_{jtr})$	-0.84	-0.84	-0.84	-0.83	-0.83	-0.83	-0.81	-0.81	-0.81
$CV(NIP_{jt})$	0.09	0.09	0.09	0.08	0.08	0.09	0.08	0.08	0.09
$CV(\epsilon_{jt})$	0.18	0.18	0.18	0.20	0.21	0.21	0.27	0.27	0.27
$CV(m_{jt})$	0.34	0.34	0.34	0.38	0.38	0.38	0.44	0.44	0.44
$\rho(T_{jtr}, H_{jt})$	0.02	0.02	0.02	0.04	0.04	0.04	0.06	0.06	0.06
$\rho(D_{jtr}, NIP_{jt})$	-0.03	-0.03	-0.03	-0.06	-0.05	-0.05	-0.05	-0.04	-0.04
$\rho(NIP_{jtr}, m_{jt})$	0.69	0.68	0.65	0.57	0.56	0.54	0.39	0.39	0.37
$CV(X_{jtr})$	1.78	1.78	1.79	1.86	1.87	1.87	2.01	2.01	2.02
Panel (d): $\alpha = -9$									
$CV(H_{jt})$	0.16	0.16	0.17	0.19	0.19	0.20	0.25	0.26	0.26
$CV(T_{jtr})$	4.74	4.75	4.76	4.34	4.35	4.37	4.33	4.35	4.37
$CV(D_{jtr})$	-0.59	-0.59	-0.59	-0.57	-0.57	-0.58	-0.57	-0.57	-0.57
$CV(NIP_{jt})$	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
$CV(\epsilon_{jt})$	0.29	0.30	0.30	0.29	0.29	0.29	0.29	0.29	0.29
$CV(m_{jt})$	0.45	0.46	0.46	0.45	0.46	0.47	0.49	0.50	0.51
$\rho(T_{jtr}, H_{jt})$	0.01	0.00	0.00	0.02	0.02	0.02	0.04	0.04	0.04
$\rho(D_{jtr}, NIP_{jt})$	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.03	-0.03	-0.03
$\rho(NIP_{jtr}, m_{jt})$	0.42	0.43	0.41	0.48	0.48	0.46	0.44	0.43	0.41
$CV(X_{jtr})$	3.59	3.59	3.59	3.73	3.74	3.76	3.91	3.92	3.94

Note: Deterministic components were drawn once. The row  $CV(X_{jtr})$  contains the mean statistics over 5,000 replications.

identified  $N^2 - N = 22,350$  importer-specific fixed effects by imposing the restrictions  $\chi_{jT} = 0$  and  $\theta_{it} = 0$  on the full set of effects<sup>16</sup>.

The MC results will be summarized by the mean bias of the estimate from the true value given in percent of the true value by  $[1/S \sum_s (\gamma - \hat{\gamma}_s)] / \gamma$  and the standard deviation of the estimate normalized by the true value  $\left[ \sqrt{1/S \sum_s (\hat{\gamma}_s - 1/S \sum_s \hat{\gamma}_s)^2} \right] / \gamma$  for the coefficient estimates on  $D_{ijt}$  and  $NIP_{jt}$ , covering only replications in which the estimation algorithms of the respective method converged. The number of successful convergences in percent of total replications is given by  $CR$  if convergence failures were present.

### 3.3 | Simulation results

Table 3 shows the normalized (in percent of true value) bias and standard deviation as well as the convergence ratio (percentage of successful convergences) of MC simulations on the DGP based on the empirical data. Panel (a) depicts the results on the full data set, that is, including intra national trade flows. The fixed effects estimator (FE) shows no bias up to the third digit in all sample sizes for the discriminatory ( $D_{ijt}$ ) as well as the nondiscriminatory ( $NIP_{jt}$ ) trade cost variable. The standard deviation on both estimates decreases with increasing  $T$  (read table from left to right), but remains constant over increasing  $N$  (read table from top to bottom).

The two-stage fixed effects estimator (FE-2S) employed on the full sample shows a small but consistent bias of between 1% and 1.2% in the discriminatory trade cost variable that is estimated in the first stage. This bias vanishes when only data on international trade flows—see Panel (b)—is utilized, since the nondiscriminatory variable, which is omitted in the first stage of this estimator, exerts a different impact on domestic and international trade flows. Hence, the first stage leads to biased estimates on the time-varying discriminatory variable, if domestic trade flows are included. In both the full and reduced sample, the second stage produces inconsistent estimates on the NDTP measures that are biased downwards by between 70% and 104%.

Contrary to the FE-2S, the BV method produces biased estimates for the coefficients on both the discriminatory and nondiscriminatory trade cost measure and both the full and reduced dataset. The bias on  $NIP_{jt}$  in Panel (a) is substantially smaller than for the FE-2S, however, reducing the sample to intra national observations the reverse is true for most of the sample sizes. Finally, the naive gravity specification yields biased and inconsistent estimates on both variables and all experiments in Table 3.

Next, we employ the empirical dataset to test the robustness to a different distributional assumption of variance of the GLM, the Gamma distribution. The results are summarized in Table 4. Most notably in the results on the full dataset given in Panel (a), the estimates on  $NIP_{jt}$  are biased for all estimators and sample sizes. In the small samples with either  $N = 50$  or  $T = 3$  the BV estimates are least biased, however, similar to the estimates of the FE-2S, the bias increases with  $N$  and  $T$ . Only the FE estimates show decreasing bias and standard deviation with increasing  $T$ . The reduction in bias progresses only slowly, that is, reducing from a bias of 25% for  $N = 150$  and  $T = 5$  to a bias of 23% for  $N = 150$  and  $T = 8$ . Also note that increasing number of countries  $N$ —and thus including smaller countries that introduce more variation—has only little and ambiguous impact on the bias. Compared with the estimators on the dataset excluding domestic observations, shown in Panel (b), the FE estimator produces the smallest bias in percentage of the true parameter values.

As a next step, we analyze if the correlation structure between the nondiscriminatory trade policy variable  $NIP_{jt}$  the latent inward multilateral resistance term  $m_{jt}$  causes the large bias in the FE-2S and BV estimators. The following simulations are based on the fully simulated DGP described above<sup>17</sup>. Table 5 summarizes the results of this scenario. Again, only the FE estimates on  $NIP_{jt}$  are unbiased

**TABLE 3** MC results: Baseline scenario (empirical dataset)

	$T = 3$						$T = 5$						$T = 8$					
	$D_{ijt}$		$NIP_{jt}$		CR	SD	$D_{ijt}$		$NIP_{jt}$		CR	SD	$D_{ijt}$		$NIP_{jt}$		CR	SD
	Bias	SD	Bias	SD			Bias	SD	Bias	SD			Bias	SD	Bias	SD		
Panel (a): full sample $N^2 \times T$																		
$N = 50$	FE	0.000	0.004	0.000	0.006	1	0.000	0.002	0.000	0.004	1	0.000	0.001	0.000	0.003	0.9999		
	FE-2S	-0.012	0.004	-0.717	0.007	1	0.010	0.002	-0.870	0.005	1	0.012	0.001	-1.037	0.004	1		
	BV	-0.012	0.015	0.100	0.027	1	-0.020	0.008	0.170	0.020	1	-0.034	0.005	0.281	0.016	1		
	Naive	-0.350	0.003	-1.137	0.001	1	-0.300	0.001	-1.025	0.001	1	-0.516	0.001	-1.040	0.000	1		
$N = 100$	FE	0.000	0.004	0.000	0.006	1	0.000	0.002	0.000	0.004	1	0.000	0.001	0.000	0.003	1		
	FE-2S	-0.010	0.004	-0.761	0.028	1	0.010	0.002	-0.654	0.025	1	0.012	0.001	-0.731	0.018	1		
	BV	-0.020	0.071	0.110	0.091	1	-0.076	0.046	0.224	0.072	1	-0.154	0.033	0.393	0.064	1		
	Naive	-0.342	0.003	-1.142	0.000	1	-0.306	0.001	-1.039	0.000	1	-0.510	0.001	-1.053	0.000	1		
$N = 150$	FE	0.000	0.004	0.000	0.006	1	0.000	0.002	0.000	0.004	1	0.000	0.001	0.000	0.003	1		
	FE-2S	-0.010	0.004	-0.763	0.028	1	0.010	0.002	-0.657	0.025	1	0.010	0.001	-0.735	0.018	1		
	BV	-0.026	0.223	0.203	0.655	1	-0.072	0.153	0.298	0.494	1	-0.164	0.078	0.496	0.464	1		
	Naive	-0.342	0.003	-1.142	0.000	1	-0.306	0.001	-1.040	0.000	1	-0.510	0.001	-1.053	0.000	1		
Panel (b): reduced sample $(N^2 - N) \times T$ , i.e., $i \neq j$																		
$N = 50$	FE-2S	0.000	0.004	-0.767	0.007	1	0.000	0.002	-0.918	0.005	1	0.000	0.001	-1.016	0.004	1		
	BV	-0.016	0.015	-1.338	0.291	1	-0.028	0.008	-1.759	0.237	1	-0.046	0.005	-3.765	0.211	1		
	Naive	-0.388	0.003	-1.208	0.000	1	-0.318	0.001	-1.056	0.000	1	-0.538	0.001	-1.041	0.000	1		

(Continues)

TABLE 3 Continued

	T = 3						T = 5						T = 8					
	$D_{ijt}$			$NIP_{jt}$			$D_{ijt}$			$NIP_{jt}$			$D_{ijt}$			$NIP_{jt}$		
	Bias	SD	CR	Bias	SD	CR	Bias	SD	CR	Bias	SD	CR	Bias	SD	CR	Bias	SD	CR
N = 100	FE-2S	0.000	0.004	-0.784	0.029	1	0.000	0.002	-0.703	0.026	1	0.000	0.001	-0.711	0.018	1		
	BV	-0.022	0.070	-1.535	2.123	1	-0.086	0.045	-1.560	2.198	1	-0.166	0.032	-3.885	1.672	1		
	Naive	-0.378	0.003	-1.209	0.000	1	-0.326	0.001	-1.072	0.000	1	-0.534	0.001	-1.057	0.000	1		
N = 150	FE-2S	0.000	0.004	-0.783	0.028	1	0.000	0.002	-0.703	0.026	1	0.000	0.001	-0.714	0.018	1		
	BV	-0.028	0.221	-1.485	11.223	1	-0.078	0.150	0.286	9.783	1	-0.170	0.076	-1.419	8.654	1		
	Naive	-0.376	0.003	-1.209	0.000	1	-0.326	0.001	-1.073	0.000	1	-0.534	0.001	-1.058	0.000	1		

<sup>a</sup>Note. All estimations are performed by PPML. Results are based on 5,000 replications.

**TABLE 4** MC results: Gamma  $\sigma_{ij}^2 = \ln(2)$  (empirical dataset)

	T = 3			T = 5			T = 8									
	D <sub>ijt</sub>			D <sub>ijt</sub>			D <sub>ijt</sub>									
	Bias	SD	CR	Bias	SD	CR	Bias	SD	CR							
Panel (a): full sample $N^2 \times T$																
N = 50	FE	0.050	0.806	-0.383	4.153	0.9996	0.022	0.434	-0.280	2.207	1	0.000	0.263	-0.239	1.575	1
	FE-2S	0.052	0.813	-0.636	0.497	1	0.032	0.432	-0.835	0.398	1	0.010	0.261	-0.995	0.443	1
	BV	-0.030	0.283	0.012	0.526	1	-0.036	0.157	0.083	0.402	1	-0.054	0.112	0.196	0.335	1
	Naive	-0.370	0.528	-1.213	0.377	1	-0.468	0.318	-1.430	0.205	1	-0.676	0.185	-1.425	0.149	1
N = 100	FE	0.034	0.743	-0.314	4.159	0.9984	0.018	0.420	-0.269	2.198	1	0.000	0.242	-0.245	1.561	1
	FE-2S	0.034	0.752	-0.669	0.546	1	0.026	0.418	-0.616	0.432	1	0.010	0.241	-0.702	0.438	1
	BV	-0.028	0.154	0.062	0.445	1	-0.090	0.095	0.189	0.327	1	-0.170	0.056	0.368	0.269	1
	Naive	-0.370	0.492	-1.217	0.368	1	-0.476	0.307	-1.441	0.189	1	-0.668	0.171	-1.429	0.138	1
N = 150	FE	0.054	0.761	-0.524	4.208	0.9994	0.006	0.416	-0.254	2.169	1	0.006	0.239	-0.226	1.549	1
	FE-2S	0.058	0.770	-0.661	0.558	1	0.014	0.414	-0.612	0.434	1	0.016	0.238	-0.698	0.432	1
	BV	-0.034	0.115	0.119	0.400	1	-0.080	0.074	0.262	0.273	1	-0.178	0.042	0.489	0.229	1
	Naive	-0.354	0.507	-1.229	0.359	1	-0.486	0.306	-1.435	0.190	1	-0.666	0.167	-1.429	0.142	1
Panel (b): reduced sample $(N^2 - N) \times T$ , i.e., $i \neq j$																
N = 50	FE-2S	0.062	0.835	-0.653	0.503	1	0.020	0.433	-0.857	0.403	1	-0.004	0.261	-0.967	0.460	1
	BV	-0.036	0.284	-1.133	7.930	1	-0.046	0.157	-1.898	6.385	1	-0.068	0.111	-4.563	5.830	1
	Naive	-0.396	0.525	-1.269	0.247	1	-0.504	0.306	-1.495	0.113	1	-0.714	0.180	-1.455	0.099	1

(Continues)

TABLE 4 Continued

	T = 3			T = 5			T = 8								
	D <sub>ijt</sub>			D <sub>ijt</sub>			D <sub>ijt</sub>								
	Bias	SD	CR	Bias	SD	CR	Bias	SD	CR						
N = 100	NIP <sub>jt</sub>			NIP <sub>jt</sub>			NIP <sub>jt</sub>								
	Bias	SD	CR	Bias	SD	CR	Bias	SD	CR						
	FE-2S	0.046	0.768	-0.667	0.554	1	0.012	0.414	0.430	1	-0.002	0.240	0.678	0.452	1
BV	-0.028	0.155	-1.394	12.950	1	-0.098	0.094	-2.178	10.417	1	-0.180	0.055	-5.357	9.727	1
Naive	-0.394	0.490	-1.282	0.227	1	-0.514	0.290	-1.503	0.106	1	-0.704	0.168	-1.454	0.093	1
N = 150	NIP <sub>jt</sub>			NIP <sub>jt</sub>			NIP <sub>jt</sub>								
	Bias	SD	CR	Bias	SD	CR	Bias	SD	CR						
	FE-2S	0.060	0.787	-0.662	0.565	1	0.006	0.412	-0.638	0.432	1	0.002	0.238	-0.674	0.442
BV	-0.034	0.115	-1.554	17.613	1	-0.086	0.074	-0.889	13.938	1	-0.184	0.042	-3.698	12.493	1
Naive	-0.380	0.505	-1.277	0.227	1	-0.518	0.290	-1.502	0.103	1	-0.700	0.165	-1.452	0.093	1

Note. All estimations are performed by PPML. Results are based on 5,000 replications.



**TABLE 5** MC results: No correlation scenario— $\rho(NIP_{jt}, m_{jt}) = 0$  (simulated dataset)

		T = 3						T = 5						T = 8							
		$D_{ijt}$			$NIP_{jt}$			$D_{ijt}$			$NIP_{jt}$			$D_{ijt}$			$NIP_{jt}$				
		Bias	SD	CR	Bias	SD	CR	Bias	SD	CR	Bias	SD	CR	Bias	SD	CR	Bias	SD	CR		
Panel (a): full sample $N^2 \times T$																					
N = 50	FE	0.002	0.117	0.002	0.121	1	0.000	0.057	0.001	0.058	1	0.000	0.033	0.000	0.032	1	0.000	0.033	0.000	0.032	1
	FE-2S	-0.956	0.023	-0.919	0.028	1	-0.970	0.011	-0.954	0.014	1	-1.012	0.005	-1.032	0.006	1	-1.012	0.005	-1.032	0.006	1
	BV	-0.448	0.099	-0.443	0.103	1	-0.528	0.048	-0.523	0.048	1	-0.632	0.026	-0.618	0.025	1	-0.632	0.026	-0.618	0.025	1
	Naive	-0.920	0.011	-0.921	0.011	1	-0.934	0.005	-0.934	0.005	1	-0.948	0.002	-0.948	0.002	1	-0.948	0.002	-0.948	0.002	1
N = 100	FE	0.002	0.058	0.003	0.058	1	0.000	0.031	0.001	0.031	1	0.000	0.019	0.000	0.018	1	0.000	0.019	0.000	0.018	1
	FE-2S	-0.980	0.009	-0.996	0.008	1	-0.986	0.004	-0.992	0.004	1	-1.016	0.002	-1.014	0.002	1	-1.016	0.002	-1.014	0.002	1
	BV	-0.438	0.049	-0.436	0.050	1	-0.542	0.025	-0.538	0.025	1	-0.694	0.014	-0.679	0.014	1	-0.694	0.014	-0.679	0.014	1
	Naive	-0.922	0.005	-0.922	0.005	1	-0.944	0.002	-0.944	0.002	1	-0.940	0.002	-0.942	0.002	1	-0.940	0.002	-0.942	0.002	1
N = 150	FE	0.000	0.039	0.000	0.038	1	0.000	0.022	0.000	0.022	1	0.000	0.013	0.000	0.013	1	0.000	0.013	0.000	0.013	1
	FE-2S	-0.998	0.006	-1.011	0.005	1	-1.000	0.003	-1.013	0.002	1	-1.010	0.001	-1.009	0.001	1	-1.010	0.001	-1.009	0.001	1
	BV	-0.440	0.034	-0.433	0.034	1	-0.574	0.018	-0.568	0.017	1	-0.730	0.010	-0.720	0.010	1	-0.730	0.010	-0.720	0.010	1
	Naive	-0.928	0.003	-0.929	0.003	1	-0.950	0.002	-0.950	0.002	1	-0.946	0.001	-0.947	0.001	1	-0.946	0.001	-0.947	0.001	1
Panel (b): reduced sample $(N^2 - N) \times T$ , i.e., $i \neq j$																					
N = 50	FE-2S	0.002	0.144	-0.078	0.073	1	0.000	0.077	0.054	0.041	1	0.000	0.046	0.610	0.309	1	0.000	0.046	0.610	0.309	1
	BV	-0.522	0.115	-0.530	0.122	1	-0.666	0.058	-0.655	0.062	1	-0.802	0.032	-0.798	0.033	1	-0.802	0.032	-0.798	0.033	1
	Naive	-0.942	0.011	-0.940	0.010	1	-0.958	0.005	-0.956	0.005	1	-0.968	0.002	-0.968	0.002	1	-0.968	0.002	-0.968	0.002	1

(Continues)

TABLE 5 Continued

	<i>T</i> = 3						<i>T</i> = 5						<i>T</i> = 8					
	<i>D<sub>ijt</sub></i>		<i>NIP<sub>jt</sub></i>		CR	CR	<i>D<sub>ijt</sub></i>		<i>NIP<sub>jt</sub></i>		CR	CR	<i>D<sub>ijt</sub></i>		<i>NIP<sub>jt</sub></i>		CR	CR
	Bias	SD	Bias	SD			Bias	SD	Bias	SD			Bias	SD	Bias	SD		
<i>N</i> = 100																		
	FE-2S	0.002	0.064	0.223	0.386	1	0.000	0.036	0.077	0.156	1	0.000	0.021	0.007	0.087	0.9922		
	BV	-0.514	0.053	-0.534	0.058	1	-0.660	0.028	-0.671	0.031	1	-0.808	0.015	-0.811	0.016	1		
	Naive	-0.934	0.005	-0.934	0.005	1	-0.954	0.002	-0.954	0.002	1	-0.950	0.002	-0.951	0.002	1		
<i>N</i> = 150																		
	FE-2S	0.000	0.041	0.219	0.066	0.999	0.000	0.024	-0.133	0.055	0.9998	0.000	0.014	-0.140	0.019	0.9994		
	BV	-0.478	0.036	-0.539	0.042	1	-0.630	0.019	-0.656	0.022	1	-0.786	0.010	-0.795	0.011	1		
	Naive	-0.936	0.003	-0.936	0.003	1	-0.956	0.002	-0.956	0.002	1	-0.950	0.001	-0.952	0.001	1		

Note. All estimations are performed by PPML. Results are based on 5,000 replications.

with decreasing bias in  $T$  and decreasing standard deviation with increasing  $N$  and  $T$ . The MC results illustrate that applying the BV or FE-2S method leads to biased and inconsistent results, even if the latent and the observed importer–time component are uncorrelated. While the bias on  $NIP_{jt}$  is smallest for the FE-2S when applied to the reduced sample, a substantial bias remains even in larger samples and there is no monotonic decrease in bias when increasing the number of periods. These findings are especially illuminating, since a violation of the fixed effect assumption could have been expected to improve the performance of this estimator. The results of this MC experiment, however, indicate that  $\rho(NIP_{jt}, m_{jt}) = 0$  is not a sufficient condition for unbiasedness of the FE-2S method.

Since the BV method is originally based on a log-linear model, Table 6 shows the results for the BV method estimated via GLM and OLS assuming a homoskedastic error, that is, setting  $\sigma_{\eta,ijt}^2 = 0.3$ . Panels (a) and (b) summarize the MC results for the baseline scenario with correlation between  $NIP_{jt}$  and the latent multilateral resistance term  $m_{jt}$  for the full and reduced sample respectively, and Panels (c) and (d) depict the results for the no correlation scenario,  $\rho(NIP_{jt}, m_{jt}) = 0$ . In the baseline scenario, the estimates of the log-linear estimator BV-LogLin on the full sample are only slightly biased, but this bias does not systematically decrease with either increasing the number of countries or periods, see Panel (a). In the no-correlation scenario, bias and standard errors decrease quickly with increasing sample size and the estimates on  $NIP_{jt}$  are virtually unbiased, as shown in Panel (c). However, excluding intra national observations in the no-correlation scenario in Panel (d), a bias of around 3% remains in the largest sample size.

Finally, the MC results on the two scenarios that increase the dispersion in trade costs and flows are shown in Table 7 for the FE estimator, along with the baseline results for comparison. Increasing the dispersion in endowments and trade costs has little effect on the properties of the FE estimator compared with the baseline scenario. Standard errors of the estimates decrease with increasing sample size along both dimensions and bias on the coefficient of  $NIP_{jt}$  strictly decreases with increasing  $T$ . We obtain similar results with somewhat smaller standard errors for the scenario with a higher elasticity of substitution  $\alpha = -9$ .

Table 8 summarizes the results of the estimations. The estimates for the MFN in the full sample vary between  $-4$  and  $-20$ , all being statistically significant on conventional levels. The naive gravity model results in estimates very similar to the BV, while the FE-2S shows estimates that are well below the values found in the empirical literature. The FE estimator results in an estimate of roughly  $-9$  which corresponds with the estimates obtained in Heid et al. (2017), who used a similar dataset. The FE, naive and 2S-FE arrive at nearly similar direct effects of regional trade agreements with coefficients between 0.14 and 0.17. The BV method, using the transformed trade cost variables, yields small, negative, and a statistically insignificant parameter estimate on  $RTA$ .

Since, the MC results indicated that the PPML might be biased when the stochastic component of the true DGP corresponds to a GLM with Gamma variance function, we additionally estimated a Gamma GLM model. This estimation is constrained by the unavailability of statistical software routines<sup>18</sup> for the Gamma PML estimator that accounts for high-dimensional fixed effects. Therefore we employed the high-dimensional fixed effect GLM routine *feglm* (see Stammann, 2018) which is only defined for strictly positive dependent variables for the Gamma family. Results for this Gamma GLM (GGLM) estimator are summarized in column (6). To enable comparison, column (5) shows the results of the PPML estimator on the same sample. Dropping the zero flow observations has a negligible impact on the PPML estimates. The coefficient on  $RTA$  is very similar for the PPML and the GGLM estimator, but the effect of the MFN decreases in size from  $-9$  to  $-6.8$ .

Panel (b) shows the results when intra national trade flows and the according observations are not available to the researcher. In this case, the FE is not applicable since it identifies the nondiscriminatory trade cost variable along that dimension of the data. The resulting estimates of the naive gravity

**TABLE 6** MC results: Homoskedastic error  $\sigma_{\eta,ijt}^2 = 0.3$  (simulated dataset)

	T = 3			T = 5			T = 8						
	D <sub>ijt</sub>		NIP <sub>jt</sub>	D <sub>ijt</sub>		NIP <sub>jt</sub>	D <sub>ijt</sub>		NIP <sub>jt</sub>				
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD			
Panel (a): full sample N <sup>2</sup> × T													
N = 50	BV-GLM	-1.418	0.750	0.085	10.636	-1.390	0.341	0.044	5.128	-1.428	0.169	0.072	2.639
	BV-LogLin	0.004	0.456	0.066	7.661	0.000	0.223	-0.009	3.551	0.002	0.113	0.012	1.775
N = 100	BV-GLM	-1.308	0.376	0.144	7.891	-1.334	0.167	-0.119	3.752	-1.346	0.085	-0.061	1.950
	BV-LogLin	-0.002	0.221	0.129	5.244	0.000	0.111	0.036	2.406	0.000	0.056	0.018	1.205
N = 150	BV-GLM	-1.332	0.250	0.214	7.289	-1.350	0.113	0.217	3.481	-1.374	0.055	0.223	1.730
	BV-LogLin	0.002	0.145	-0.022	4.214	0.000	0.071	0.019	1.958	0.000	0.037	-0.024	0.985
Panel (b): reduced sample (N <sup>2</sup> -N) × T, i.e., i ≠ j													
N = 50	BV-GLM	-1.436	0.740	3.366	84.851	-1.404	0.340	-2.975	50.072	-1.436	0.169	-3.918	28.157
	BV-LogLin	0.004	0.463	-8.227	223.249	0.002	0.225	-6.552	129.879	0.002	0.114	2.324	71.893
N = 100	BV-GLM	-1.326	0.367	1.620	69.670	-1.350	0.165	2.714	38.991	-1.358	0.084	0.157	17.569
	BV-LogLin	-0.002	0.222	-15.848	240.955	0.000	0.111	-8.049	132.386	0.000	0.056	-6.825	81.977
N = 150	BV-GLM	-1.344	0.249	-1.476	68.713	-1.358	0.112	-1.000	0.000	-1.382	0.055	-1.000	0.000
	BV-LogLin	0.002	0.145	-11.175	244.451	0.000	0.071	-19.339	136.693	0.000	0.037	-8.355	82.315

(Continues)

TABLE 6 Continued

	T = 3			T = 5			T = 8					
	$D_{ijt}$			$D_{ijt}$			$D_{ijt}$					
	Bias	SD	$NIP_{jt}$	Bias	SD	$NIP_{jt}$	Bias	SD	$NIP_{jt}$			
Panel (c): $\rho(NIP_{jt}, m_{jt}) = 0$ , full sample $N^2 \times T$												
N = 50	BV-GLM	-0.468	0.382	-0.463	0.389	0.192	-0.533	0.194	-0.642	0.105	-0.630	0.102
	BV-LogLin	-0.008	0.238	-0.008	0.242	0.122	0.001	0.123	0.000	0.068	-0.001	0.068
N = 100	BV-GLM	-0.450	0.222	-0.447	0.226	0.113	-0.551	0.115	-0.700	0.060	-0.686	0.059
	BV-LogLin	0.000	0.123	0.001	0.124	0.064	0.000	0.064	0.000	0.036	0.001	0.036
N = 150	BV-GLM	-0.446	0.157	-0.440	0.159	0.078	-0.574	0.079	-0.736	0.042	-0.726	0.041
	BV-LogLin	0.000	0.080	-0.001	0.082	0.043	0.002	0.044	0.000	0.024	0.000	0.024
Panel (d): $\rho(NIP_{jt}, m_{jt}) = 0$ , reduced sample $(N^2 - N) \times T$ , i.e., $i \neq j$												
N = 50	BV-GLM	-0.540	0.443	-0.554	0.510	0.226	-0.661	0.256	-0.806	0.126	-0.802	0.140
	BV-LogLin	-0.012	0.265	-0.070	0.727	0.140	-0.027	0.347	-0.006	0.080	-0.035	0.179
N = 100	BV-GLM	-0.522	0.230	-0.540	0.267	0.120	-0.681	0.137	-0.812	0.064	-0.815	0.071
	BV-LogLin	0.000	0.128	-0.045	0.492	0.068	-0.022	0.239	0.000	0.039	-0.024	0.110
N = 150	BV-GLM	-0.484	0.158	-0.545	0.211	0.082	-0.659	0.103	-0.788	0.042	-0.798	0.051
	BV-LogLin	-0.002	0.083	-0.106	0.511	0.045	-0.055	0.236	0.000	0.025	-0.033	0.108

Note. All estimations denoted by "BV-GLM" are performed by PPM, and those denoted by "BV-LogLin" via OLS. Results are based on 5,000 replications.

**TABLE 7** MC results: Different scenarios for FE estimator (simulated dataset)

	$N$	$T = 3$			$T = 5$			$T = 8$						
		$D_{ijt}$			$NIP_{jt}$			$NIP_{jt}$						
		Bias	SD	$D_{ijt}$	Bias	SD	$D_{ijt}$	Bias	SD	$D_{ijt}$	Bias	SD	$NIP_{jt}$	Bias
(a) Baseline	50	0.002	0.333	-0.019	1.913	0.000	0.160	-0.015	0.902	-0.002	0.082	0.001	0.473	
	100	0.006	0.181	-0.001	1.284	0.004	0.088	-0.006	0.617	0.000	0.045	-0.003	0.312	
	150	0.000	0.133	0.011	1.188	0.000	0.064	0.009	0.563	0.000	0.032	-0.002	0.290	
(b) $v_z = 0.3$	50	-0.002	0.394	-0.010	1.355	0.002	0.187	0.011	0.641	0.000	0.093	0.003	0.340	
	100	0.002	0.234	0.002	1.166	-0.002	0.109	0.002	0.558	0.000	0.054	-0.001	0.286	
	150	0.000	0.180	-0.007	1.080	-0.002	0.084	-0.005	0.514	0.000	0.042	0.000	0.263	
(c) $\alpha = -9$	50	0.008	0.324	-0.010	1.052	-0.004	0.161	0.003	0.526	0.002	0.082	0.001	0.293	
	100	0.000	0.176	0.006	0.725	0.000	0.084	0.005	0.342	0.000	0.044	-0.002	0.186	
	150	0.000	0.125	-0.006	0.652	0.000	0.060	-0.001	0.311	0.000	0.031	0.000	0.167	

Note. All estimations are performed by PPML. Results are based on 5,000 replications.



**TABLE 8** Estimation results: Empirical application

	(1) naive	(2) BV	(3) FE-2S	(4) FE	(5) FE <sup>1</sup>	(6) FE GGLM <sup>1</sup>
Panel (a): full sample $N^2 \times T$						
$\ln(1 + MFN_{jt})$	-3.96***	-4.36**	-20.43**	-8.97***	-8.99***	-6.80***
(0.423)	(1.65)	(7.91)	(0.87)	(0.87)	(0.84)	
$RTA_{ijt}$	0.14***	-0.05	0.19***	0.14***	0.13***	0.15***
	(0.03)	(0.18)	(0.03)	(0.03)	(0.03)	(0.03)
Included fixed effects						
$\theta_{ij}$	Yes	Yes	Yes	Yes	Yes	Yes
$\lambda_{it}$	No	No	Yes	Yes	Yes	Yes
$\chi_{jt}$	No	No	Yes	Yes	Yes	Yes
Observations	21,060	21,060	21,060/320	21,060	20,579	20,579
Panel (b): reduced sample $(N^2 - N) \times T$ , i.e., $i \neq j$						
$\ln(1 + MFN_{jt})$	-4.15***	-161.13*	1.02	—	—	—
	(0.42)	(69.51)	(5.05)	—	—	—
$RTA_{ijt}$	0.14***	-0.18	0.07**	—	—	—
	(0.03)	(0.10)	(0.03)	—	—	—
Included fixed effects						
$\theta_{ij}$	Yes	Yes	Yes	—	—	—
$\lambda_{it}$	No	No	Yes	—	—	—
$\chi_{jt}$	No	No	Yes	—	—	—
Observations	20,735	20,735	20,735/319	—	—	—

Notes. All estimations are performed by PPML except for column (6), which is estimated by Gamma GML. Heteroskedasticity robust standard errors in parentheses. \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ . <sup>1</sup>Estimated on  $X_{ijt} > 0$ .

model are very similar to those obtained in the full sample. By contrast, the coefficient on the MFN of the BV-method drastically decreases to -161 accompanied by a loss in precision. The estimate on the RTA remains insignificant and negative. The FE-2S method yields a statistically insignificant positive coefficient of around 1 on the MFN. Moreover, the coefficient on regional trade agreements declines from 0.19 in the full to a less precisely estimated 0.07 in the reduced sample.

Overall, the results of this empirical exercise demonstrate that the BV and FE-2S approaches result in implausible coefficients on either the discriminatory, the nondiscriminatory, or both trade cost variables. Comparing the results between the full and reduced sample size, the naive gravity model results are more robust than the results obtained from these two methods. In this specific application, disregarding the zero flow observations had only a minor impact on the FE estimates. When interpreted as a trade elasticity, the Gamma GML estimate on the most-favored nation tariff would be closer to the values given in Head and Mayer (2014) than the PPML estimate.

## 4 | EMPIRICAL APPLICATION

This section illustrates the methods discussed above in an empirical application. It should be noted that it is not the purpose of this application to derive exact estimates on the effects of the trade cost variables used, but to demonstrate the impact of the methods on the estimates of the NDTP variable. For a thorough analysis of the effects of trade policy measures on trade flows in structural gravity model, the challenges summarized in Piermartini and Yotov (2016) should be adequately addressed. Hence, we employ a panel dataset including intra national observations of 3-year intervals, apply a PPML and control for directional pairwise fixed effects.

The following application is based on the dataset used in Oberhofer and Pfaffermayr (2017) augmented by data on an importer-based NDTP variable. The dataset covers 65 countries between 2000 and 2012 in 3-year intervals, that is, five periods. Trade flows cover an aggregate of manufacturing industries and are based on OECD-STAN, UNIDO's INDSTAT database, CEPII and WIOD—Oberhofer and Pfaffermayr (2017), for a more detailed description of the dataset. In this dataset trade flows are normalized by total world production at time  $t$  such that  $\sum_i \sum_j X_{ijt} = 1$ . The respective production and expenditures are given by  $PROD_{it} = \sum_j X_{ijt}$  and  $EXP_{jt} = \sum_i X_{ijt}$ . The dataset also includes information on the formation of regional trade agreements that is taken from “Mario Larch's Regional Trade Agreements Database” (see Egger & Larch, 2008). As a measure of nondiscriminatory trade policy the most-favored nation (MFN) tariff was selected. We use the simple average of the applied tariff rate over all products obtained from UNCTAD's Trade Analysis Information System (TRAINS) accessed via World Integrated Trade Solution (WITS). Data was cross-checked with the WTO Tariff Reports, and the most plausible value was taken<sup>19</sup>. Missing data values for some periods were linearly interpolated between the most recent values before and after the period of the missing value. As is common practice with data on tariffs, the variable will be included as  $\ln(MFN_{ij} + 1)$ . The descriptive statistics of the data are summarized in Table A1 in the Appendix.

Prior to the estimation we removed all observations of directional country-pairs that have missing or recorded zero flows over all periods, since these observations cannot be identified by the directional pairwise fixed effects. From 21,125 observations, a total of 65 are removed according to this procedure, leaving us with 21,060 observations. Estimations are performed on the full sample including intra national observations and on a reduced sample covering only the with  $21,060 - (65 \cdot 5) = 20,735$  international observations. The estimators for the (1) naive, (2) BV, (3) FE-2S, and (4) FE are applied as described in Section 2, including all transformations of the data. Hence, the FE is estimated using the interaction with the international border dummy  $NIP_{jt} \times INT_{ij}$ .

Table 8 summarizes the results of the estimations. The estimates for the MFN in the full sample vary between  $-4$  and  $-20$ , all being statistically significant on conventional levels. The naive gravity model results in estimates very similar to the BV, while the FE-2S shows estimates that are well below the values found in the empirical literature. The FE estimator results in an estimate of roughly  $-9$  which corresponds to the estimates obtained in Heid et al. (2017), who used a similar dataset. The FE, naive and 2S-FE arrive at nearly similar direct effects of regional trade agreements with coefficients between 0.14 and 0.17. The BV method, using the transformed trade cost variables, yields small, negative, and a statistically insignificant parameter estimate on  $RTA$ .

Since, the MC results indicated that the PPML might be biased when the stochastic component of the true DGP corresponds to a GLM with Gamma variance function, we additionally estimated a Gamma GLM model. This estimation is constrained by the unavailability of statistical software routines<sup>20</sup> for Gamma PML estimator that accounts for high-dimensional fixed effects. Therefore we employed the high-dimensional fixed effect GLM routine *feglm* (see Stammann, 2018) which is only

defined for strictly positive dependent variables for the Gamma family. Results for this Gamma GLM (GGLM) estimator are summarized in column (6). To enable comparison, column (5) shows the results of the PPML estimator on the same sample. Dropping the zero flow observations has a negligible impact on the PPML estimates. The coefficient on RTA is very similar for the PPML and the GGLM estimator, but the effect of the MFN decreases in size from  $-9$  to  $-6.8$ .

Panel (b) shows the results when intra national trade flows and the according observations are not available to the researcher. In that case, the FE is not applicable since it identifies the nondiscriminatory trade cost variable along that dimension of the data. The resulting estimates of the naive gravity model are very similar to those obtained in the full sample. By contrast, the coefficient on the MFN of the BV-method drastically decreases to  $-161$  accompanied by a loss in precision. The estimate on the RTA remains insignificant and negative. The FE-2S method yields a statistically insignificant positive coefficient of around 1 on the MFN. Moreover, the coefficient on regional trade agreements declines from 0.19 in the full to a less precisely estimated 0.07 in the reduced sample.

Overall, the results of this empirical exercise demonstrate that the BV and FE-2S approaches result in implausible coefficients on either the discriminatory, the nondiscriminatory or both trade cost variables. Comparing the results between the full and reduced sample size, the naive gravity model results are more robust than the results obtained from these two methods. In this specific application, disregarding the zero flow observations had only a minor impact on the FE estimates. When interpreted as a trade elasticity, the Gamma GML estimate on the most-favored nation tariff would be closer to the values given in Head and Mayer (2014) than the PPML estimate.

## 5 | DISCUSSION AND CONCLUSIONS

In this paper, Monte Carlo (MC) evidence on the consistency and bias of estimators for identifying the effects of nondiscriminatory trade policy (NDTP) variables on trade flows in a structural gravity model (SGM) were presented. The benchmark included the fixed effects estimator (FE) on data covering intra national observations (as suggested by Heid et al., 2017), and two estimators that have frequently been used in empirical research on datasets without intra national observations: the bonus-vetus (BV) method (see Baier & Bergstrand, 2009) and a two-stage fixed effects (FE-2S) procedure (as suggested by Head & Mayer, 2014; Piermartini & Yotov, 2016). The latter two methods have been used or suggested to overcome the collinearity problem that researchers face when trying to identify NDTP effects while using exporter–time and importer–time fixed effects or other theory-consistent estimation frameworks.

The data generating process (DGP) is based on an economic motivation of the SGM that satisfies the structural restrictions that are postulated by theory and addresses the empirically observed heteroskedasticity of the data. We constructed these processes by utilizing empirical panel data on GDP and trade costs, and by simulating a dataset to allow for fine-tuning of the underlying data properties. Following the suggestions for structural gravity modeling in Piermartini and Yotov (2016), the DGP and all estimators take into account time-invariant directional pairwise trade costs. The bias and standard deviation of Poisson pseudo-maximum likelihood (PPML) estimates on the NDTP measure of the different econometric estimators have then been analyzed in a series of MC scenarios on sample sizes varying in both  $N$  and  $T$ .

The results demonstrated that the FE estimator, which is only identified for datasets including intra national observations, yields unbiased and consistent estimates in all scenarios that assume an error distribution that is efficiently estimated via PPML. Reductions in bias are driven by increasing the number of periods  $T$  in the panel, while the standard errors around the mean estimates decrease with

increasing  $N$  as well. Results of the MC scenario with an error distribution corresponding to Gamma-GLM indicated biased estimates on the NDTP parameter even for the largest sample size with  $N=150$  and  $T=8$ . However, the bias is again decreasing in  $T$  and standard errors decrease in both  $N$  and  $T$ .

Results of the baseline scenario indicate strongly biased and inconsistent estimates on the NDTP variable obtained by the BV and FE-2S on samples including or excluding intra national observations. Assuming homoskedastic errors, we illustrated that the bias of the BV method by PPML is due to a combination of functional form misspecification, correlation between the latent multilateral resistance term (MRT) and the NDTP variable, and missing intra national flow observations. Since the assumption of no correlation between the MRT and NDTP variable will not hold in practice, the BV method can be expected to produce biased coefficient estimates on NDTP variables. The FE-2S method produces persistently biased and inconsistent estimates, independent of assumptions regarding the correlation between the MRT and the NDTP variable. This finding is particularly revealing, since one may have expected the two-stage FE procedure to produce unbiased results given the fixed effects assumption is violated.

Three recommendations for practitioners may be derived from the results of this paper. First and foremost, assuming that NDTP impact only border-crossing trade flows, we recommend to use the approach outlined Heid et al. (2017) for econometric estimation of a theory-consistent SGM. The MC results on bias and consistency showed, that this estimator has favorable asymptotic properties under very general assumptions. In contrast, the BV and FE-2S methods that have been frequently employed in past empirical research, often because domestic trade flows were unavailable, yield biased and inconsistent point estimates. Given a DGP, similar to the one analyzed in this paper, researchers should refrain from ad hoc solutions to tackle collinearity between NDTP variables of interest and fixed effects that capture the structure of the model. The results of the empirical application emphasize this finding. Gathering data on intra national trade flows introduces an additional variation that can be exploited for proper identification, while still conforming to the structure of the theoretical model. Following this recommendation may prove to be challenging, since readily available datasets on intra national trade are still scarce and can have substantial publication lags. Alternatively, the domestic trade flows maybe constructed by the researcher using trade flows and national gross production statistics as, for example, in Oberhofer and Pfaffermayr (2017) or Heid et al. (2017).

Secondly, the MC results showed that the favorable asymptotic properties of the FE in terms of bias on the NDTP estimate work through the number of periods  $T$  included. Hence, given a trade off between the number of countries  $N$  and periods  $T$  when constructing a dataset, we recommend researchers to favor the latter over the further when interested in the effects of NDTP measures. However, maximizing the data coverage with respect to the periods included will be a particular problem when following the recommendations in Piermartini and Yotov (2016) of using 3- or 5-year intervals to allow for adjustments of trade flows to policy changes. Moving from aggregate manufacturing data to industry-level or product-group data further aggravates the problem. Moreover, analysis on this level will potentially be confronted with zero export or import flows of a product and country or on some dyadic pairs for all periods. In such cases, not all importer–period, exporter–period or directional pairwise fixed effects will be identified.

Third, the MC results illustrated that, assuming a GLM with a Gamma variance process as the true DGP, the FE estimated by PPML—while showing consistent behavior—yielded biased estimates for the NDTP variable, even for the longest panel (with  $T=8$ ) considered in the simulations. We therefore recommend researchers to accompany their Poisson PML results with a Gamma PML estimates, especially when confronted with small sample sizes. Compared to the Poisson PML, the Gamma PML puts even less weight on observations with large conditional mean (see Santos Silva & Tenreyro, 2006), which in some cases may improve the model fit. Model selection could then be based on information

criteria, specification tests or prediction error statistics (see Martínez-Zarzoso, 2013; Santos Silva & Tenreyro, 2006). Endeavors to follow this recommendation might be hindered by the availability of statistical software routines that support both a Gamma PML—and thus zero trade flows—as well as high-dimensional fixed effects.

The findings of this paper are subject to several limitations that are open to further research. The MC results are focussed on the bias and consistency of the coefficient estimates, but did not examine the efficiency of the estimators or if and how strongly the estimated standard errors are affected by the incidental parameter problem. Future research could build upon the work of Pfaffermayr (2017, 2018), who showed that his derived constrained PPML—which leads to identical point estimates as a dummy variable PPML in cases of no missing data—yields asymptotically unbiased standard errors. Furthermore, in the DGP employed in this paper, we made ad hoc assumptions on the distribution of the time-invariant directional pairwise trade costs, since there is no clear empirical benchmark for the size of those effects. Further analysis on the impact of different correlation structures or dispersion properties of these effects would be insightful. Another limitation relates to the performance when confronted with a large number of zero trade flows in the data, since the theory-consistent DGP used in this paper does not generate such flows. Hence, augmenting this DGP, with a process that specifically produces zero flows (as, e.g., in Santos Silva & Tenreyro, 2011), while still obeying the structure of the model, could provide further insights.

## NOTES

<sup>1</sup>Note that empirical research increasingly employs panel data sets and/or covers specific sectors or product groups. Exploiting the time variation via a panel data structure is strongly recommended in Piermartini and Yotov (2016), whereas they suggest to include every third or fifth year to permit adjustment of trade flows to policy changes.

<sup>2</sup>For simplicity, the exporter-specific preference parameter from the original specification of Anderson and Van Wincoop (2003) is omitted.

<sup>3</sup>Various types of models motivated from the demand or supply side suggested in the literature yield a model of the type presented here (see, e.g., Anderson & Van Wincoop, 2003; Chaney, 2008; Eaton & Kortum, 2002; Melitz, 2003). These models only differ in their interpretation of the elasticity of trade flows with respect to trade costs  $\alpha$ .

<sup>4</sup>The structural part can be obtained by assuming that Equation 1 only holds in expectation, replacing  $X_{ijt}$  by  $E(X_{ijt})$ .

<sup>5</sup>Figueiredo, Lima and Orefice (2016) argued that if one assumes that the log-linear model is identified, then the GLM estimates will be severely biased. Since the researcher cannot know which model is true, they proposed a robust quantile estimation approach.

<sup>6</sup>Given  $i \neq l$  and  $j \neq k$  there are numerous sets making estimation computationally demanding. To speed up the computation, the estimation may be performed on only a subset of all possible combinations, leading to efficiency losses.

<sup>7</sup>The dummy variable FE approach will in particular be computationally less demanding when optimization methods for handling high-dimensional fixed effects are used (as, e.g., Larch, Wanner, Yotov, & Zylkin, 2018).

<sup>8</sup>Available datasets include CEPII's "TradeProd" (see De Sousa, Mayer, & Zignago, 2012), that covers approximately 150 countries over the years 1980 to 2006 at a three-digit level of the ISIC Revision 2 manufacturing industries and the World Input Output Database (see Timmer, Dietzenbacher, Los, Stehrer, & De Vries, 2015) covering 43 countries over the years 2000 to 2014 at 56 ISIC Rev. 4 industries.

<sup>9</sup>Note that assuming a log-linear model and that all bilateral variables are exogenous, the HTM will yield estimates similar to the second stage of the fixed effects vector decomposition and the two-stage FE model.

<sup>10</sup>While this approach is appealing, adoption of the Head and Mayer (2014) DGP in a GLM framework with heteroskedastic errors is not straight forward.

<sup>11</sup>We chose the largest 150 countries in terms of GDP, for which data on GDP was completely available between 1994 and 2015.

- <sup>12</sup>A total of 384 values were interpolated this way.
- <sup>13</sup>Since the most-favored nation tariff is a direct price shifter, the parameter can be interpreted as the trade elasticity or elasticity of substitution.
- <sup>14</sup>Note that Head and Mayer (2014) advised against the use of a negative binomial GLM, because of the scale dependency of its estimates, as shown in Bosquet and Boulhol (2014). Thus we restrict our analysis to a Gamma variance function.
- <sup>15</sup>First, orthogonalize a standard normal variable ( $z_{ji}$ ) to the centered and scaled effect  $m_{ji}$ , then scale both vectors to length one ( $\bar{m}_{ji}$ ,  $\bar{z}_{ji}$ ) and compute  $\tilde{z}_{ji} = \bar{z}_{ji} + 1/\tan(\arccos(r)) * \bar{m}_{ji}$  for an exact sample correlation  $r$ , with  $r = 0$ .
- <sup>16</sup>For the two-stage fixed effects estimator on the reduced sample excluding intra national observations, we additionally impose the identifying restrictions  $\theta_{Nj} = \theta_{iN} = \theta_{N-1, N-2} = 0$ .
- <sup>17</sup>The baseline results of the simulated DGP are summarized in Table A2 in the Appendix. Compared with the baseline results of the empirical DGP, the parameter estimates on  $NIP_{ji}$  of the naive, BV and FE-2S estimators are even more biased, but the main results regarding the asymptotic properties are similar.
- <sup>18</sup>The *glm* routine in Stata for a Gamma family with a logarithmic link function and the full set of fixed effects did not converge.
- <sup>19</sup>For example, Switzerland is reported to have a most-favored nation tariff of zero for all products according to the TRAINS database. However, the WTO Tariff Report show a positive value for the simple average of the MFN tariff.
- <sup>20</sup>The *glm* routine in Stata for a Gamma family with a logarithmic link function and the full set of fixed effects did not converge.

## REFERENCES

- Anderson, J. E. & Van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, 93 (1), 170–192.
- Anderson, J. E. & Yotov, Y. V. (2016). Terms of trade and global efficiency effects of free trade agreements, 1990–2002. *Journal of International Economics*, 99, 279–298.
- Baier, S. L. & Bergstrand, J. H. (2009). Bonus vetus OLS: A simple method for approximating international trade-cost effects using the gravity equation. *Journal of International Economics*, 77(1), 77–85.
- Baier, S. L. & Bergstrand, J. H. (2010). Approximating general equilibrium impacts of trade liberalizations using the gravity equation. In P. Van Bergeijk & S. Brakman (Eds.), *The gravity model in international trade: Advances and applications* (Chap. 4, pp. 88–134). Cambridge, U.K.: Cambridge University Press.
- Bergstrand, J. H., Larch, M., & Yotov, Y. V. (2015). Economic integration agreements, border effects, and distance elasticities in the gravity equation. *European Economic Review*, 78, 307–327.
- Beverelli, C., Neumueller, S., & Teh, R. (2015). Export diversification effects of the WTO trade facilitation agreement. *World Development*, 76, 293–310.
- Bosquet, C. & Boulhol, H. (2014). Applying the GLM variance assumption to overcome the scale-dependence of the negative binomial QGPML estimator. *Econometric Reviews*, 33(7), 772–784.
- Bratt, M. (2017). Estimating the bilateral impact of nontariff measure on trade. *Review of International Economics*, 25(2), 1–25.
- Burger, M., Van Oort, F., & Linders, G.-J. (2009). On the specification of the gravity model of trade: Zeros, excess zeros and zero-inated estimation. *Spatial Economic Analysis*, 4(2), 167–190.
- Cameron, A. C. & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge, U.K.: Cambridge University Press.
- Chaney, T. (2008). Distorted gravity: The intensive and extensive margins of international trade. *American Economic Review*, 98(4), 1707–1721.
- Charbonneau, K. B. (2012). *Multiple fixed effects in nonlinear panel data models—theory and evidence (unpublished manuscript)*. Princeton, NJ: Princeton University.
- De Sousa, J., Mayer, T., & Zignago, S. (2012). Market access in global and regional trade. *Regional Science and Urban Economics*, 42(6), 1037–1052.
- Eaton, J. & Kortum, J. E. S. (2001). Trade in capital goods. *European Economic Review*, 45(7), 1195–1235.
- Eaton, J. & Kortum, S. (2002). Technology, geography and trade. *Econometrica*, 70(5), 1741–1779.



- Egger, P. (2005). Alternative techniques for estimation of cross-section gravity models. *Review of International Economics*, 13(5), 881–891.
- Egger, P., Francois, J., Manchin, M., & Nelson, D. (2015). Non-tariff barriers, integration and the transatlantic economy. *Economic Policy*, 30(83), 539–584.
- Egger, P. & Larch, M. (2008). Interdependent preferential trade agreement memberships: An empirical analysis. *Journal of International Economics*, 76(2), 384–399.
- Egger, P. & Staub, K. (2016). GLM estimation of trade gravity models with fixed effects. *Empirical Economics*, 50, 137–175.
- Feenstra, R. (2004). *Advanced international trade: Theory and evidence*. Princeton, NJ: Princeton University Press.
- Fernández-Vala, I. & Weidner, M. (2016). Individual and time effects in nonlinear panel models with large N, T. *Journal of Econometrics*, 192(1), 291–312.
- Figueiredo, E., Lima, L. R., & Orefice, G. (2016). Migration and regional trade agreements: A (new) gravity estimation. *Review of International Economics*, 24(1), 99–125.
- Greene, W. (2011). Fixed effects vector decomposition: A magical solution to the problem of time-invariant variables in fixed effects models? *Political Analysis*, 19(2), 135–146.
- Gylfason, T., Martínez-Zarzoso, I., & Wijkman, P. M. (2015). Free trade agreements, institutions and the exports of eastern partnership countries. *Journal of Common Market Studies*, 53(6), 1214–1229.
- Harrigan, J. (1996). Openness to trade in manufacturers in the OECD. *Journal of International Economics*, 40(1–2), 23–39.
- Hausman, J. A. & Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica*, 49(6), 1377–1398.
- Head, K. & Mayer, T. (2014). Gravity equations: Workhorse, toolkit, and cookbook. In G. Gopinath, E. Helpman, & K. Rogoff (Eds.), *Handbook of international economics* (Vol. 4, pp. 131–195). Amsterdam, The Netherlands: Elsevier.
- Head, K., Mayer, T., & Ries, J. (2009). How remote is the offshoring threat? *European Economic Review*, 53(4), 429–444.
- Head, K. & Ries, J. (2008). FDI as an outcome of the market for corporate control: Theory and evidence. *Journal of International Economics*, 74(1), 2–20.
- Heid, B., Larch, M., & Yotov, Y. V. (2017). *Estimating the effects of nondiscriminatory trade policies within structural gravity models* (CESifo Working Paper No. 6735). Munich, Germany: CESifo.
- Hoekman, B. & Nicita, A. (2011). Trade policy, trade costs, and developing country trade. *World Development*, 39(1), 2069–2079.
- Jochmans, K. (2017). Two-way models for gravity. *Review of Economics and Statistics*, 99(3), 478–485.
- Larch, M., Wanner, J., Yotov, Y. V., & Zylkin, T. (2018). Currency union and trade: A PPML re-assessment with high-dimensional fixed effects. *Oxford Bulletin of Economics and Statistics*. <https://doi.org/10.1111/obes.12283>
- Lederman, D., Olarreaga, M., & Payton, L. (2010). Export promotion agencies: Do they work? *Journal of Development Economics*, 91(2), 257–265.
- Lee, H. & Park, I. (2007). In Search of optimised regional trade agreements and applications to east Asia. *The World Economy*, 30(5), 783–806.
- Martínez-Zarzoso, I. (2013). The log of gravity revisited. *Applied Economics*, 45(3), 311–327.
- McCallum, J. (1995). National borders matter: Canada–U.S. regional trade patterns. *American Economic Review*, 85(3), 615–623.
- Melitz, J. M. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6), 1695–1725.
- Melitz, J. M. (2008). Language and foreign trade. *European Economic Review*, 52(4), 667–699.
- Moisé, E., Orlic, T., & Minor, P. (2011). *Trade facilitation indicators: The potential impacts on developing countries trade* (OECD Trade Policy Paper No. 118). Paris, France: OECD Publishing.
- Moisé, E. & Sorescu, S. (2013). *Trade facilitation indicators: The potential impact of trade facilitation on developing countries' trade* (OECD Trade Policy Papers, No. 144). Paris, France: OECD Publishing, Paris.
- Oberhofer, H. & Pfaffermayr, M. (2017). *Estimating the trade and welfare effects of brexit: A panel data structural gravity model* (CESifo Working Papers). Munich, Germany: CESifo.
- Oberhofer, H., Pfaffermayr, M., & Sellner, R. (2018). *Revisiting time as a trade barrier: Evidence from a panel structural gravity model* (CESifo Working Paper No. 7426). Munich, Germany: CESifo.
- Pfaffermayr, M. (2017). *Constrained Poisson pseudo maximum likelihood estimation of structural gravity models* (eecon Working Paper). Innsbruck, Austria: University of Innsbruck.

- Pfaffermayr, M. (2018). *Trade creation and trade diversion of regional trade agreements revisited: A constrained panel pseudo-maximum likelihood approach (eeecon Working Papers in Economics and Statistics No. 2018-10)*. Innsbruck, Austria: University of Innsbruck.
- Piermartini, R. & Yotov, Y. V. (2016). *Estimating trade policy effects with structural gravity (CESifo Working Paper No. 6009)*. Munich, Germany: CESifo.
- Plümper, T. & Troeger, V. E. (2007). Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects. *Political Analysis*, 15(2), 124–139.
- Portugal-Perez, A. & Wilson, J. S. (2012). Export performance and trade facilitation reform: Hard and soft infrastructure. *World Development*, 40(7), 1295–1307.
- Santos Silva, J. M. C. & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, 88 (4), 641–658.
- Santos Silva, J. M. C. & Tenreyro, S. (2011). Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator. *Economics Letters*, 112(2), 220–222.
- Stammann, A. (2018). *Fast and feasible estimation of generalized linear models with high-dimensional k-way fixed effects (ArXiv e-prints)*. Ithaca, NY: Cornell University.
- Timmer, M. P., Dietzenbacher, E., Los, B., Stehrer, R., & De Vries, G. J. (2015). An illustrated user guide to the world input-output database: The case of global automotive production. *Review of International Economics*, 23, 575–605.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Sellner R. Non-discriminatory Trade Policies in Panel Structural Gravity Models: Evidence from Monte Carlo Simulations. *Rev Int Econ*. 2019;00:1–34.  
<https://doi.org/10.1111/roie.12401>